Econometric Theory, **18**, 2002, 349–386. Printed in the United States of America.
DOI: 10.1017.S0266466602182065

ROBUST ESTIMATION OF STRUCTURAL BREAK POINTS

INMACULADA FITENI
Universidad Carlos III de Madrid

This paper proposes robust M-estimators of dynamic linear models with a structural break of unknown location. Rates of convergence and limiting distributions for the estimated shift point and the estimated regression parameters are derived. The analysis is carried out in the framework of possibly dependent observations and also with trending regressors. The asymptotic distribution of the break location estimator is obtained both for fixed magnitude of shift and for shift with magnitude converging to zero as the sample size increases. The latter is essential for the derivation of feasible confidence intervals for the break location. Monte Carlo simulations illustrate the performance of asymptotic inferences in practice.

1. INTRODUCTION

The development of valid statistical inference tools in the presence of a structural break with unknown location has been a major concern in the statistic and econometric literature. Among the survey papers on this topic we mention Deshayes and Picard (1986), Krishnaiah and Miao (1988), Csörgo and Horváth (1988), Hušková and Sen (1989), Hušková (1997), and Stock (1997). There are also several monographs, such as Broemeling and Tsurumi (1986), Hackl (1989), Hackl and Westlund (1989, 1991), and Brodsky and Darkhovsky (1993). A complete review related to this problem can be found in Csörgo and Horváth (1997).

The main objectives are to test whether a change in the model parameters has occurred and, if so, to estimate its location and magnitude. Testing for the presence of a structural break is a research topic with a long-standing tradition. Particularly important contributions include the CUSUM test of Brown, Durbin and Evans (1975) and Hackl (1980) and its robust versions in Hušková (1990), the Wald, Lagrange multiplier, and likelihood-ratio-like tests of Andrews (1993), the exponentially weighted tests of Andrews and Ploberger (1994),

This article is part of my doctoral dissertation at Universidad Carlos III de Madrid. I thank Donald W.K. Andrews and three anonymous referees for very useful comments that considerably improved the quality of this paper. I am grateful to my thesis advisers, Miguel A. Delgado and Javier Hidalgo, for their support and suggestions. The research was funded by a postdoctoral fellowship from Comunidad de Madrid and European Social Fund, no. 4323/2000, by Dirección General de Enseñanza Superior, no. PB98-0025, and by Ministerio de Ciencia y Tecnología, BEC2001-1270. Address correspondence to: Inmaculada Fiteni; Departamento de Estadística y Econometría; Universidad Carlos III de Madrid; C/ Madrid, 126; 28903 Madrid; e-mail: fiteni@est-econ.uc3m.es.

 0266-4666/02 $9.50

and the predictive tests of Ghysels, Guay, and Hall (1997), among others. A number of studies concerning the issue of multiple structural changes are emerging; see, e.g., Andrews, Lee, and Ploberger (1996), Garcia and Perron (1996), Liu, Wu, and Zidek (1997), and Bai and Perron (1998).

The literature addressing the issue of structural change points estimation is comparatively sparse, however. First developments include the maximum likelihood estimation of the break date with a simple shift. This is considered by Hinkley (1970), Bhattacharya (1987), and Yao (1987) for the independent and identically distributed (i.i.d.) case, by Picard (1985) for a Gaussian autoregressive process, and by Feder (1975) for segmented regressions. Bai (1994) estimates the unknown change point by least squares, considering a linear process for the error term. Bai and Perron (1998) generalize this result allowing for multiple shifts in the regression model. But these classical estimators are sensitive to deviations from the model distribution, to outlying observations, and to model misspecifications, which can produce disastrous effects on the estimates. Departures from the assumed model can be solved, in part, by estimating nonparametrically the underlying regression model, as proposed by Carlstein (1988), Dümbgen (1991), Chu and Wu (1993), and Delgado and Hidalgo (2000), among others. Alternatively, robust methods, which are insensitive to small deviations from the assumptions, can also be applied. Bai (1995) proposes using the LAD estimator, which has good properties in terms of robustness (qualitative robustness, B-robustness, and maximum breakdown point). However, this estimation procedure is highly inefficient under normality. Hušková (1997) proposes an M-estimator for the location of a change in the mean of a sequence of i.i.d. observations, obtaining the best trade-off between efficiency under the true model and robustness under a possibly thick-tailed distribution.

In this paper, we consider a linear regression model with structural change and propose the M-estimators of both the regression parameters and the shift location, which jointly optimize a robust objective function. Most of the estimators can be obtained as particular cases of this one. To perform inferences, our main goal is to derive the rates of convergence and the limiting distributions of the estimators, which will be obtained both for fixed and for shrinking magnitude of shift. The latter is essential for the derivation of feasible confidence intervals for the break point location, provided that only in this case the asymptotic distribution will not be case dependent. These asymptotic results also hold for the scale invariant version of the estimator, but rather restrictive assumptions will be required, as in the standard context of no change by Yohai (1987). General forms of serial dependence and also trending regressors will be allowed.

The rest of the paper is organized as follows. Section 2 introduces the model and regularity conditions. The asymptotic properties of the M-estimators are studied in Section 3, under the two assumptions on the magnitude of shift, fixed and converging to zero. In Section 4, the finite sample performance of the asymptotic approximation is illustrated by means of a small Monte Carlo experiment.

2. THE MODEL AND ASSUMPTIONS

Let $\{Z_t = (Y_t, X_t)\}_{t=1}^{n}$ be a sample of Z, a $\mathbb{R} \times \mathbb{R}^p$-valued stochastic process defined on the probability space $(\Omega, \mathcal{F}, P)$, such that

$$Y_t = X_t'\beta_{10} I(t \leq [n\tau_0]) + X_t'\beta_{20} I(t > [n\tau_0]) + U_t, \tag{1}$$

a linear model with a single shift, where $I(A)$ is the indicator function of the event A, $[\cdot]$ represents the nearest integer function, and $\{U_t\}_{t=1}^{n}$ is the sequence of disturbance terms. Here $\beta_{j0} \in \Theta \subset \mathbb{R}^p$, for $j = 1$ and 2, are the unknown regression parameters for each regime and $\tau_0 \in \Pi \subset (0,1)$ represents the shift point location, which is also unknown. The size of the jump will be denoted by $\lambda = \beta_{10} - \beta_{20}$.

Under the maintained hypothesis that the shift exists, that is, $\lambda \neq 0$,[1] we propose to estimate the unknown parameter vector $\xi_0 = (\beta_{10}', \beta_{20}', \tau_0)'$, which will be defined by the following assumption.

A0.

$$\xi_0 = \underset{\xi \in \Theta^2 \times \Pi}{\arg\min} \lim_{n \to \infty} E[S_n(\xi)], \tag{2}$$

with $S_n(\xi) = S_{1n}(\beta_1, \tau) + S_{2n}(\beta_2, \tau)$, such that

$$S_{1n}(\beta, \tau) = \frac{1}{n} \sum_{t=1}^{[n\tau]} \rho_\sigma(Y_t - X_t'\beta) \quad \text{and}$$

$$S_{2n}(\beta, \tau) = \frac{1}{n} \sum_{t=[n\tau]+1}^{n} \rho_\sigma(Y_t - X_t'\beta), \tag{3}$$

where $\rho_\sigma(u) = \rho(u/\sigma)$, $\rho : \mathbb{R} \to \mathbb{R}$ is a function that identifies the model parameters, and σ is such that $\lim_{n\to\infty} n^{-1} \sum_{t=1}^{n} E[\rho(U_t/\sigma)] = d$, for $d = E_\Phi[\rho(u)]$ and Φ represents the standard normal distribution.

Remark 1. Given the model (1), the "identification assumption" A0 is not satisfied for every ρ-function. At the end of this section, we provide the adequate setup for A0 to hold. As will be shown in Proposition 2, the requirements assumed to obtain the robust and asymptotic properties for the estimators will assure the fulfillment of A0.

Remark 2. Each ρ-function defines a particular linear predictor of Y_t given X_t, with changing parameters at a given moment of time $[n\tau_0]$. For instance, $\rho(u) = u^2$ defines the least squares predictor, $\rho(u) = |u|$ the least absolute deviation predictor, and $\rho(u) = \frac{1}{2}u^2 I(|u| \leq c) + |u| I(|u| > c)$ is the Huber predictor, a compromise between the previous two, given a suitable constant c. Different ρ-functions define, in general, different parameter values though, in certain circumstances, they can be identical. For instance, when the conditional distribution of Y_t given X_t is symmetric with respect to its mean, which is a linear combination of the X_t with changing parameters, least squares (LS), least absolute deviations (LAD), and Huber predictors have identical values. However, the resulting estimators will have different statistical properties.

A natural estimate of ξ_0 will be given by

$$\hat{\xi}_n = \underset{\xi \in \Theta^2 \times \Pi}{\arg\min}\ \mathcal{S}_n(\xi), \tag{4}$$

the sample analogue of (2), which can be obtained by implementing an iterative procedure. First, for each $\tau \in \Pi$, we obtain the regression parameter estimators pre-$[n\tau]$ and post-$[n\tau]$ as

$$\hat{\beta}_{jn}(\tau) = \underset{\beta \in \Theta}{\arg\min}\ \mathcal{S}_{jn}(\beta, \tau), \tag{5}$$

for $j = 1$ and 2, respectively. Second, the shift point will be estimated as the sample partition that minimizes the objective function concentrated in τ, i.e.,

$$\hat{\tau}_n = \underset{\tau \in \Pi}{\arg\min}(\mathcal{S}_{1n}(\hat{\beta}_{1n}(\tau), \tau) + \mathcal{S}_{2n}(\hat{\beta}_{2n}(\tau), \tau)). \tag{6}$$

Thus, $\hat{\xi}_n = (\hat{\beta}'_{1n}, \hat{\beta}'_{2n}, \hat{\tau}_n)'$, where $\hat{\beta}_{jn} = \hat{\beta}_{jn}(\hat{\tau}_n)$ are the coefficient estimators for $j = 1$ and 2. The size of the jump will be estimated by $\hat{\lambda}_n = \hat{\beta}_{1n} - \hat{\beta}_{2n}$.

Remark 3. To obtain a scale invariant estimator, we must consider the objective function in (4) satisfying (3) with $\rho_\sigma(u)$ replaced by $\rho_{\hat{s}_n}(u)$, where $\hat{s}_n$ is a consistent estimator of σ, which can be obtained either separately or simultaneously with $\hat{\xi}_n$. This estimator is called "*M*-estimate with general scale." A consistent and robust estimator of the scale can be obtained from a preliminary consistent estimator $\tilde{\xi}_n$ of ξ_0, such that

$$\sum_{t=1}^{[n\tilde{\tau}_n]} \chi\left(\frac{Y_t - X_t'\tilde{\beta}_{1n}}{\hat{s}_n}\right) + \sum_{t=[n\tilde{\tau}_n]+1}^{n} \chi\left(\frac{Y_t - X_t'\tilde{\beta}_{2n}}{\hat{s}_n}\right) = 0,$$

for a given function $\chi : \mathbb{R} \to \mathbb{R}$. For example, the median absolute deviation (MAD) is defined by $\chi(u) = sign(|u| - 1)$, or Huber's proposal is given by $\chi(u) = \psi^2(u) - b$, for $b = E_\Phi[\psi^2(u)]$, where $\psi(u) = \partial\rho(u)/\partial u$ is the score function. If we consider different scales for each subsample, the estimator could be defined as $\hat{s}_n = \hat{s}_{1n} I(t \le [n\tilde{\tau}_n]) + \hat{s}_{2n}(t > [n\tilde{\tau}_n])$, where

$$\sum_{t=1}^{[n\tilde{\tau}_n]} \chi\left(\frac{Y_t - X_t'\tilde{\beta}_{1n}}{\hat{s}_{1n}}\right) = 0 \quad \text{and} \quad \sum_{t=[n\tilde{\tau}_n]+1}^{n} \chi\left(\frac{Y_t - X_t'\tilde{\beta}_{2n}}{\hat{s}_{2n}}\right) = 0.$$

In Corollary 2, we provide the conditions under which the coefficient estimators obtained previously will be asymptotically equivalent to those obtained with an assumed known scale for the error term. Henceforth, for simplicity and without loss of generality, we will consider the objective function given by (3) and (4) for $\sigma = 1$, such that $\rho_\sigma(u) = \rho(u)$.

As a matter of notation, we let $\|\cdot\|$ denote the Euclidean norm of a vector or a matrix and $\|\cdot\|_r$ the L_r norm of a random q-vector (i.e., $\|X\|_r = (\sum_{i=1}^{q} E|X_i|^r)^{1/r}$). The symbol $\rightarrow^p$ represents convergence in probability, $\rightarrow^d$ convergence in distribution, $\stackrel{d}{=}$ equivalence in distribution of events and $\Rightarrow$ weak convergence in the space $D[0,1]$ under the Skorokhod metric (see, e.g., Pollard, 1984).

Next, we define the concept of near epoch dependence (NED), the type of asymptotically weak dependent structure assumed for the data. A NED process will be "approximately" mixing in the sense of being well approximated by the near epoch of a mixing process and includes linear processes, strong mixing processes, and many other dependent structures as special cases (see, e.g., Davidson, 1994). Under suitable conditions, it will have properties permitting the application of limit theorems, of which the mixingale property is the most important. This idea was introduced by Ibramigov (1962), and has been formalized in different ways by Billingsley (1968), McLeisch (1975a, 1975b), Bierens (1981), Andrews (1988), Wooldridge and White (1988), Hansen (1991), and Pötscher and Prucha (1991), among others.

DEFINITION 1. *Let* $\{V_t\}_{-\infty}^{\infty}$ *be a strong mixing sequence, possibly vector-valued, on a probability space* $(\Omega, \mathcal{G}, P)$ *and define* $\mathcal{G}_{t-m}{}^{t+m} = \sigma(V_{t-m}, \dots, V_{t+m})$ *such that* $\{\mathcal{G}_{t-m}^{t+m}\}_{m=0}^{\infty}$ *is an increasing sequence of* σ*-subfields of* $\mathcal{G}$*. For* $r \geq 0$*, a sequence of integrable random vectors* $\{W_t\}_{-\infty}^{\infty}$ *is said to be* L_r*-NED of size* $-q_0$ *on the strong mixing base* $\{V_t\}$ *of size* $-q_1$ *if there exists a sequence of nonnegative constants* $\{d_t\}_1^{\infty}$ *and a nonnegative sequence* $\{v_m\}_0^{\infty}$*, such that* $v_m \rightarrow 0$ *as* $m \rightarrow 0$*, and,*

(i) *for* $r = 0$, $\Pr(\|W_t - E[W_t|\mathcal{G}_{t-m}^{t+m}]\| > \varepsilon) \leq d_t v_m, \quad \forall \varepsilon > 0,$
(ii) *for* $r > 0$, $\|W_t - E[W_t|\mathcal{G}_{t-m}^{t+m}]\|_r \leq d_t v_m,$

hold for all $t \geq 1$ *and* $m \geq 0$*. Besides,* $v_m = O(m^{-q})$ *for all* $q > q_0$ *and* $\{\alpha_m\}_{m\geq 0}$*, the sequence of the strong mixing numbers of* $\{V_t\}$*, is such that* $\alpha_m = O(m^{-q})$ *for all* $q > q_1$*.*

Observe that an L_q-NED sequence will be L_p-NED for $1 \leq p \leq q$ by Liapunov's inequality and for $0 = p \leq q$ by Markov's. The statistical properties of the resulting estimators are obtained in the next section under the following set of assumptions A1 and A2.

A1. Assumptions on $\rho(\cdot)$.

A1.1. Let $\rho : \mathbb{R} \rightarrow \mathbb{R}$ be a convex real function, twice continuously differentiable, with first derivative ψ and such that (i) $\lim_{u\rightarrow\pm\infty} \rho(u) = \infty$, (ii) $\rho(u) = \rho(-u)$, and (iii) $\rho(u)$ is nondecreasing for $u \geq 0$.

A1.2. $\psi : \mathbb{R} \rightarrow \mathbb{R}$ is a bounded function, and there exists a constant $m > 0$ such that $\psi(u) = a > 0$ for $u > m$.

A2. Model Assumptions. Given $\theta \in \Theta$, define the sequence $\{\eta_t(\theta)\}_{t\le n}$, where $\eta_t(\theta) = \psi(U_t + \theta' X_t)X_t$ for each t. This sequence arises from the first-order conditions, which outline the M-estimator of the regression coefficients, evaluated in a neighborhood $\theta' X_t$ of the error term, given the model (1). Let $\eta_t = \eta_t(0_p)$, $\forall t \le n$, where 0_p is a p-vector of zero. The subscript t of these sequences indicates the dependency on the data $\{Z_t\}$, and θ could be dependent on n, in which case it will be denoted by θ_n.

A2.1. $\Theta \subset \mathbb{R}^p$ is a compact and convex set.
A2.2. $\tau_0 \in \Pi$, provided that Π has closure in (0,1).
A2.3. $\{Z_t = (Y_t, X_t')'\}_{t\le n}$ is a random vector with domain in Z, L_0-*NED* on a strong mixing base $\{w_t : t = \dots,0,1,\dots\}$ with constants $d_t = 1$, where Z is a Borel subset of $\mathbb{R}^{p+1}$ defined on the probability space $(\Omega, \mathcal{F}, P)$. Let $\mathcal{F}_n = n^{-1}\sum_1^n \mathcal{F}(Z_t)$, such that $\{\mathcal{F}_n\}_{n\ge 1}$ is tight on Z.
A2.4. For some $r > 2$, $\{\eta_t\}_{t\le n}$ is a random vector sequence of mean zero, L_2-*NED* of size $-\frac{1}{2}$ on a strong mixing base $\{w_t : t = \dots,0,1,\dots\}$ of size $-r/(r-2)$, with constants $d_t = 1$ and $\sup_{t\le n} E\|\eta_t\|^r < \infty$.
A2.5. $\forall\theta \in \Theta$, $\eta_t(\theta)$ is Borel measurable in Z_t and $\dot\eta_t(\theta) = \partial\eta_t(\theta)/\partial\theta'$, continuous in $(Z_t, \theta) \in Z \times \Theta$ by A1.1, satisfies that $\sup_{t\le n} E[\sup_{\theta\in\Theta}\|\dot\eta_t(\theta)\|^{1+\varepsilon}] < \infty$, for some $\varepsilon > 0$.
A2.6. There exists a b_0 such that, for all $b > b_0$, the smallest eigenvalues of the positive matrices $b^{-1}\sum_{t=[n\tau_0]+1}^{[n\tau_0]+b} \dot\eta_t(\theta)$ and $b^{-1}\sum_{t=[n\tau_0]-b}^{[n\tau_0]} \dot\eta_t(\theta)$ are bounded away from zero uniformly in $\theta \in \Theta$.
A2.7. The $\lim_{n\to\infty} n^{-1}\sum_{t=1}^{[n\tau]} E[\dot\eta_t(\theta)]$ exists uniformly in $(\theta,\tau) \in \Theta \times \Pi$ and equals $\tau M(\theta)$ $\forall(\theta,\tau) \in \Theta \times \Pi$, where $M(\theta) = \lim_{n\to\infty} n^{-1}\sum_{t=1}^{n} E[\dot\eta_t(\theta)]$ is a positive definite matrix $\forall\theta \in \Theta$. For notational convenience, define $M = M(0_p)$.
A2.8. $\forall\tau \in (0,1)$, $\lim_{n\to\infty} \mathrm{var}[n^{-1/2}\sum_{t=1}^{[n\tau]}\eta_t] = \tau S$, where $S = \lim_{n\to\infty} \times \mathrm{var}[n^{-1/2}\sum_{t=1}^{n}\eta_t]$ is a finite and positive $p \times p$ matrix.

Assumptions A1 are standard in robust estimation. The differentiability of ρ allows one to solve (5) from the first-order conditions; i.e., we obtain

$$\{\hat\beta_{1n}(\tau)\} = \left\{\beta \in \Theta : \quad \sum_{t=1}^{[n\tau]} \eta_t(\beta_{0t} - \beta) = 0\right\},$$

$$\{\hat\beta_{2n}(\tau)\} = \left\{\beta \in \Theta : \quad \sum_{t=[n\tau]+1}^{n} \eta_t(\beta_{0t} - \beta) = 0\right\},$$

as estimators of

$$\begin{aligned}\{\beta_1(\tau)\} &= \left\{\beta \in \Theta : \quad \lim_{n\to\infty} \sum_{t=1}^{[n\tau]} E[\eta_t(\beta_{0t} - \beta)] = 0\right\},\\ \{\beta_2(\tau)\} &= \left\{\beta \in \Theta : \quad \lim_{n\to\infty} \sum_{t=[n\tau]+1}^{n} E[\eta_t(\beta_{0t} - \beta)] = 0\right\},\end{aligned} \tag{7}$$

respectively, where $\beta_{0t} = \beta_{10} I(t \le [n\tau_0]) + \beta_{20} I(t > [n\tau_0])$. For each τ, $\{\hat\beta_{1n}(\tau)\}$ and $\{\hat\beta_{2n}(\tau)\}$ define those subsets of Θ where the objective function in (4) is minimized given the observed data. Their convexity is guaranteed by the convexity of ρ. Moreover, A1 assure that $\{\beta_1(\tau)\}$ and $\{\beta_2(\tau)\}$ will be non-empty, convex, and compact sets. If ρ were also strictly convex, these subsets of Θ would shrink to an unique point (Huber, 1964), and the estimation problem could be simplified.[2] But this requirement rules out estimators like Huber's, which could be of interest. The same occurs with the condition of a continuous second derivative of ρ, which we assume in A1.1. Nevertheless, in Proposition 1, which follows, we obtain a uniformly convergent smoothed version of the Huber score function, which will prevent us from excluding this type of estimator. The rest of the requirements stated in A1 guarantee robustness against heavy-tailed distributions of the error term.

The asymptotic properties of the parameter estimators are obtained assuming convexity of the parametric space, in A2.1 and A2.2. The latter also considers the shift location far away of the interval extremes. The change point is represented by $[n\tau_0]$, such that each temporal segment increases proportionately with the sample size. Assumptions A2.3 and A2.4 are standard requirements of weak dependence outlined for this robust regression context. These assumptions are general enough to allow more primitive dependence conditions on both residuals and regressors, taking into account that certain transformations of NED sequences preserve this type of dependence structure (see, e.g., Davidson, 1994, Sec. 17.3). For simplicity and without loss of generality, we set the constants $\{d_t\}$ to 1. This could be relaxed to include trending moments, and the results would be easy to obtain. Assumption A2.5 is assumed because the break point is estimated by a global search. Assumption A2.6 states that there must be enough observations near the true break point so its identification could be possible. Assumptions A2.7 and A2.8 are standard requirements to obtain covariance stationary conditions for the asymptotic distribution of the estimated parameters.

PROPOSITION 1. *Given the Huber score function, $\psi(t) = ct\min\{|t|/c,1\}/|t|$, for a suitable constant c, we obtain that the following sequence of twice differentiable functions,*

$$h_n(t) = \left\{ \begin{array}{ll} c, & t > c + \dfrac{1}{2n^p} \\ -\dfrac{n^p}{2}t^2 + n^p\left(c + \dfrac{1}{2n^p}\right)t - \left(\dfrac{n^p c^2}{2} - \dfrac{c}{2} + \dfrac{1}{8n^p}\right), & c - \dfrac{1}{2n^p} \le t \le c + \dfrac{1}{2n^p} \\ t, & |t| < c - \dfrac{1}{2n^p} \\ \dfrac{n^p}{2}t^2 + n^p\left(c + \dfrac{1}{2n^p}\right)t + \left(\dfrac{n^p c^2}{2} - \dfrac{c}{2} + \dfrac{1}{8n^p}\right), & -c \le t \le -c + \dfrac{1}{2n^p} \\ -c, & t < -c - \dfrac{1}{2n^p} \end{array} \right\},$$

converges to $\psi(t)$ uniformly in t, for $p > 0$, a fixed c, and large n.

Thus, the asymptotic results derived for smooth ψ functions can apply to Huber's as in Bloomfield and Steiger (1983, Theorem 2), where LAD estimators are considered. This is an alternative way to solve the problem with a non-differentiable objective function, which has already been considered by others such as Jurečková and Sen (1993). The latter outline this type of function as a sum of three functions with different degrees of smoothing and study the asymptotic behavior of each one. Proposition 1 can be extended to another non-differentiable functional. The proof is omitted here to save space and can be found in Fiteni (1999), available upon request.

Remark 4. As pointed out earlier, the serial dependence conditions established in A2.3 and A2.4 allow different forms of temporal dependence for the dependent variable and/or the perturbance term, subject to suitable restrictions on their moments. However, the stationarity requirements A2.7 and A2.8 rule out models with a structural change in the marginal distribution of the regressors, which can be of interest. For instance, the linear processes are excluded when the coefficients associated with the lagged dependent variables are subject to change. To see this, consider an AR(1) model with a structural break, i.e., $U_t = Y_t - \rho_1 Y_{t-1} I(t \leq [n\tau_0]) + \rho_2 Y_{t-1} I(t > [n\tau_0])$ and let $\{U_t\}$ be a scalar sequence i.i.d. $(0, \sigma_u^2)$. Note that the process $\{\eta_t\}$ will be L_2-NED on $\{U_t\}$ if $\{\psi(U_t)\}$ is L_2-bounded and $|\rho_i| < 1$ for $i = 1,2$.[3] Moreover, we have that $S = (\tau_0 \Omega_1 + (1 - \tau_0)\Omega_2)$, where $\Omega_i = E[\psi^2(U_t)]\sigma_u^2/(1 - \rho_i^2)$, for $i = 1,2$. But condition A2.8 does not hold. In fact, we have that $\lim_{n\to\infty} \operatorname{var}[n^{-1/2} \sum_{t=1}^{[n\tau]} \psi(U_t) Y_{t-1}] = (\tau\Omega_1 + (\tau - \tau_0)(\Omega_2 - \Omega_1) I(\tau > \tau_0)) \neq \tau S$; i.e., the second moments matrix of the cumulated data does not grow linearly in this case, as assumed in the paper. But, in general, this requirement could easily be relaxed (the same argument will apply for A2.7) supposing that there also exists a shift in the second moments of the regressors at the time $[n\tau_0]$ so that, for each $\tau \in \Pi$,

$$\lim_{n\to\infty} \operatorname{var}\left[n^{-1/2} \sum_{t=1}^{[n\tau]} \eta_t\right] = \tau S_1 + (\tau - \tau_0)(S_2 - S_1) I(\tau > \tau_0) = S(\tau), \tag{8}$$

with $S_1 = \lim_{n\to\infty} \operatorname{var}[(n\tau_0)^{-1/2} \sum_{t=1}^{[n\tau_0]} \eta_t]$ and $S_2 = \lim_{n\to\infty} \operatorname{var}[(n - n\tau_0)^{-1/2} \sum_{t=[n\tau_0]+1}^{n} \eta_t]$, finite and positive matrices. And the asymptotic results of Section 3, which follows, could also apply considering that $S = S(1)$.

Finally, as pointed out in Remark 1, we study in the following proposition the identification conditions of ξ_0 that assure A0; see Rothenberg (1971) regarding the identification concept.

PROPOSITION 2. *Assume (1), A1.1, A1.2, A2.1–A2.5, and A2.7; then A0 holds. In particular,*

(2.1) For each τ, the parameters $\beta_j(\tau)$, defined by (7), are obtained as

$$\beta_1(\tau) = \beta_{10} + (\tau_0 - \tau)\mathcal{M}_1^{-1} M(\tilde{\theta}_2)'\lambda I(\tau \geq \tau_0),$$

$$\beta_2(\tau) = \beta_{20} + (\tau_0 - \tau)\mathcal{M}_2^{-1} M(\tilde{\theta}_1)'\lambda I(\tau \leq \tau_0),$$

where $\mathcal{M}_1 = \tau_0 M(\tilde{\theta}_1) + (\tau - \tau_0) M(\tilde{\theta}_2)$, $\mathcal{M}_2 = (\tau_0 - \tau) M(\bar{\theta}_1) + (1 - \tau_0) M(\bar{\theta}_2)$, $\tilde{\theta}_i = \tilde{\delta}_i(\beta_1(\tau) - \beta_{i0})$, *and* $\bar{\theta}_i = \bar{\delta}_i(\beta_2(\tau) - \beta_{i0})$, *for* $i = 1,2$, *such that* $0 < \tilde{\delta}_i, \bar{\delta}_i < 1$.

(2.2) *There is no parameter vector* $(\beta_1'(\tau), \beta_2'(\tau), \tau)' \in \Theta \times \Pi$ *observationally equivalent to* ξ_0, *which will be defined by (2) in A0.*

Observe that, by (2.1) in Proposition 2, when τ equals the true change point location, the parameter vector $(\beta_1'(\tau), \beta_2'(\tau))'$ coincides with the true regression coefficients $(\beta_{10}', \beta_{20}')'$. For the LS (a particular case, with $\psi(u) = u$), it holds that $\mathcal{M}_1^{-1} M(\tilde{\theta}_2)'\lambda = \tau^{-1}\lambda$ and $\mathcal{M}_2^{-1} M(\bar{\theta}_1)'\lambda = (1 - \tau)^{-1}\lambda$, and $(\beta_1'(\tau), \beta_2'(\tau))'$ will be defined by means of certain linear combinations of the true regression parameters, with weights depending on the relative position between the points τ and τ_0.[4] The consistency property of the estimator $\hat{\xi}_n$ can be deduced straightforwardly from the preceding proposition.

3. ASYMPTOTIC PROPERTIES

To establish the asymptotic distribution of the estimators, first we need to derive their rates of convergence. These are obtained in Theorem 1, which follows.

THEOREM 1. *Under A1.1, A1.2, and A2.1–A2.7, it holds that*

$$(\hat{\beta}_{jn} - \beta_{j0}) = Op\left(\frac{1}{\sqrt{n}}\right), \quad \text{for } j = 1 \text{ and } 2,$$

$$(\hat{\tau}_n - \tau_0) = Op\left(\frac{1}{n\|\lambda\|^2}\right). \tag{9}$$

The rate of convergence is inferred from the global behavior analysis of the objective function $\mathcal{S}_n(\xi)$ over the whole parameter space. To this end, observe that the parameter estimator (4) can also be defined as

$$\hat{\xi}_n = \underset{\xi \in \Theta^2 \times \Pi}{\arg\min} \left(\mathcal{S}_n(\xi) - \mathcal{S}_n(\xi_0)\right), \tag{10}$$

and we want to prove that $\forall \varepsilon > 0$, $\exists C > 0$ such that

$$\Pr\left\{\left(\|\hat{\beta}_{1n} - \beta_{10}\| > \frac{C}{\sqrt{n}}\right) \cup \left(\|\hat{\beta}_{2n} - \beta_{20}\| > \frac{C}{\sqrt{n}}\right) \cup \left(\|\hat{\tau}_n - \tau_0\| > \frac{C}{n\|\lambda\|^2}\right)\right\} < 3\varepsilon. \tag{11}$$

The upper bound of 3ε is chosen only for notational convenience, and, without loss of generality, it corresponds to one ε for each of the three sets. By the definition of $\hat{\xi}_n$ in (10), $\mathcal{S}_n(\hat{\xi}_n) - \mathcal{S}_n(\xi_0) \le 0$, so the left side of (11) will be upper bounded by

$$\Pr\left\{\inf_{A \cup B \cup D} \left(\mathcal{S}_n(\xi) - \mathcal{S}_n(\xi_0)\right) < 0\right\} = \Pr\left\{\sup_{A \cup B \cup D} \left(\mathcal{S}_n(\xi_0) - \mathcal{S}_n(\xi)\right) > 0\right\},$$

where the sets A, B, and D are defined as follows:

$$A = \left\{\beta_1 \in \Theta \,/\, \|\beta_1 - \beta_{10}\| > \frac{C}{\sqrt{n}}\right\}, \tag{12}$$

$$B = \left\{\beta_2 \in \Theta \,/\, \|\beta_2 - \beta_{20}\| > \frac{C}{\sqrt{n}}\right\}, \tag{13}$$

$$D = \left\{\tau \in \Pi \,/\, \|\tau - \tau_0\| > \frac{C}{n\|\lambda\|^2}\right\}. \tag{14}$$

Thus, Theorem 1 is a consequence of the following result.

THEOREM 2. *Under A1.1, A1.2, and A2.1–A2.7, it holds that* $\forall \varepsilon > 0$, $\exists C > 0$ *such that*

$$\Pr\left\{\sup_{A \cup B \cup D} (\mathcal{S}_n(\xi_0) - \mathcal{S}_n(\xi)) > 0\right\} < 3\varepsilon,$$

where A, B, and D are defined in (12), (13), and (14), respectively.

The estimators of the regression coefficients are, as usual, $\sqrt{n}$-consistent. The rate of convergence corresponding to the break point estimator depends on the magnitude of the shift, in such a way that the larger the break is, the easier its identification will be. This fact will also allow one to incorporate two standard settings for λ: fixed and shrinking to zero as n increases. For the latter, it will be denoted by λ_n, such that $\|\lambda_n\| \to 0$ with $n\|\lambda_n\|^2 \to \infty$. Then, from (9), the rate of convergence of $\hat{\tau}_n$ is obtained to be $Op(n^{-1})$ when the break is constant and $Op(n^{-1}\|\lambda_n\|^{-2})$ for a local change. In both cases this estimator will be consistent, however.

As viewed previously, Theorem 2 describes the global behavior of $\mathcal{S}_n(\xi)$ on the whole parametric set $\Theta^2 \times \Pi$. In this context, the rate of convergence must always be obtained before the asymptotic distribution, given that the arg min functional used to define the location estimate is not continuous when the minimized function is defined on an unbounded set, and the continuous mapping theorem would not follow. But the limiting distribution can be obtained by studying the local behavior of this objective function on a compact set determined by Theorem 2. Thus, β_j, for $j = 1$ and 2, is constrained to be in an $n^{-1/2}$ neighborhood of the true parameter values, β_{j0}, for $j = 1$ and 2, respectively, and a similar comment applies to the shift point, in a neighborhood of τ_0 determined by its rate of convergence. This will allow us to reparametrize the objective function (10) in a suitable form as follows:

$$\Lambda_n(v) = \mathcal{S}_n(\xi_0 + (n^{-1/2}v_1', n^{-1/2}v_2', n^{-1}P_\lambda v_3)') - \mathcal{S}_n(\xi_0),$$

for $v = (v_1', v_2', v_3)' \in V_N \subset \mathbb{R}^p \times \mathbb{R}^p \times \mathbb{R}$, with $V_N = \{v : \|v_j\| < N, j = 1,2,3\}$, a compact set defined for an arbitrary constant $N > 0$. Besides, $P_\lambda = O(\|\lambda\|^{-2})$, such that we will consider $P_{\lambda_n} = O(\|\lambda_n\|^{-2})$ for the decreasing case or $P_\lambda = 1$ for a constant λ. Thus,

$$\beta_j = \beta_{j0} + n^{-1/2}v_j, \quad \text{for } j = 1 \text{ and } 2,$$
$$\tau = \tau_0 + n^{-1}P_\lambda v_3.$$

The weak convergence result follows taking into account that $\sqrt{n}(\hat\beta_{jn} - \beta_{j0}) = \hat v_j$, for $j = 1$ and 2, and $n(\hat\tau_n - \tau_0) = P_\lambda \hat v_3$, such that $(\hat v_1', \hat v_2', \hat v_3)' = \arg\min_{v\in V_N} \Lambda_n(v)$, defined on a compact set for $N < \infty$. This is obtained in Theorem 3, which follows.

THEOREM 3. *Under A1.1, A1.2, and A2.1–A2.8, it holds that*

(i) For the coefficient estimators,

$$\begin{bmatrix} \sqrt{n}(\hat\beta_{1n} - \beta_{10}) \\ \sqrt{n}(\hat\beta_{2n} - \beta_{20}) \end{bmatrix} \xrightarrow{d} M^{-1}S^{1/2} \begin{bmatrix} \tau_0^{-1/2}Z_p & 0_{p\times p} \\ 0_{p\times p} & (1-\tau_0)^{-1/2}Z_p \end{bmatrix}, \tag{15}$$

where Z_p represents the p-dimensional standard Gaussian vector and $0_{p\times p}$ is a $p \times p$-matrix of zeros.

(ii) Assuming $\lambda_n \to 0$ with $n\|\lambda_n\|^2 \to \infty$,

$$\frac{(\lambda_n' M \lambda_n)^2}{\lambda_n' S \lambda_n} n(\hat\tau_n - \tau_0) \Rightarrow \arg\max_w \left\{ W(w) - \frac{1}{2}|w| \right\}, \tag{16}$$

where $W(\cdot)$ represents an independent two-sided standard Brownian motion defined in $\mathbb{R}$.

(iii) Assuming λ constant,

$$n(\hat\tau_n - \tau_0) \Rightarrow \arg\max_w \left\{ \lambda' W^*(w) - \frac{1}{2}\lambda' M(\lambda)\lambda|w| \right\}, \tag{17}$$

where $W^*(\cdot)$ represents a process defined in $\mathbb{Z}$, the integer set, such that

$$W^*(w) = \begin{Bmatrix} 0, & w = 0 \\ \sum_{t=w}^{-1} \eta_t, & w = -1, -2, \dots \\ \sum_{t=1}^{w} \eta_t, & w = 1, 2, \dots. \end{Bmatrix} \tag{18}$$

(iv) The distribution of $\sqrt{n}((\hat\beta_{1n} - \beta_{10})', (\hat\beta_{2n} - \beta_{20})')'$ and that of $n(\hat\tau_n - \tau_0)$ are asymptotically independent for the two cases of λ.

Part (i) of Theorem 3 asserts that the estimated regression parameters have a standard limiting distribution, as if the true change point were known. For the shift estimator, we only obtain a free distribution under the assumption of λ

decreasing with n, in part (ii). This is characterized by a two-sided Brownian motion, defined as

$$W(w) = \begin{Bmatrix} W_1(-w), & w < 0 \\ W_2(w), & w \geq 0 \end{Bmatrix},$$

with $\{W_1(t): t \in [0,\infty)\}$ and $\{W_2(t): t \in [0,\infty)\}$ being two independent standard Brownian processes. The explicit form of the distribution in (16) is given by

$$F(t) = \frac{1}{2} + \frac{1}{\sqrt{2\pi}}\sqrt{t}e^{-(1/8)t} + \frac{3}{2}e^{t}\Phi\left(-\frac{3}{2}\sqrt{t}\right) - \left(\frac{1}{2}t + \frac{5}{2}\right)\Phi\left(-\frac{1}{2}\sqrt{t}\right), \quad \textbf{(19)}$$

for $t > 0$. See, e.g., Bai (1994) or Hušková (1997) and references therein. It can be easily seen that for this local change case, the asymptotic variance of the shift estimator depends on the ratio

$$\frac{(\lambda_n' M \lambda_n)^2}{(\lambda_n' S \lambda_n)} = \frac{(\lambda_n' M(\mathcal{F}, \psi)\lambda_n)^2}{(\lambda_n' S(\mathcal{F}, \psi)\lambda_n)},$$

which again depends on the distribution $\mathcal{F}$, defined in A2.3, and the score ψ function, in A1.2. The larger the ratio is, the smaller the asymptotic variance will be. Thus, for this regression case, the smallest variance corresponds to $\psi(U_t(\theta)) = f(U_t(\theta))^{-1}\partial f(U_t(\theta))/\partial\theta'$, where $U_t(\theta) = (Y_t - \theta' X_t)$, $\theta \in \Theta$, and $f(\cdot)$ denotes the density function of $U_t(\theta)$, inferred from the distribution $\mathcal{F}(Z_t)$. If this density were unknown, we could develop an estimator of the optimal score function that could be used as the proper score function. Part (iii) provides the limiting distribution of the break point estimator under a fixed magnitude of shift. If the right-hand side of (17) were a set with more than one element, the continuous mapping theorem would not hold. But, as in Bai (1995), the problem could be modified as follows. We redefine $\hat{\tau}_n$ as the smallest value of the set of those elements holding (6). Then, $n(\hat{\tau}_n - \tau_0)$ will converge weakly to the distribution corresponding to the minimum value of the set determined by the distribution in (17), which is uniquely defined. However, this functional central limit theorem is useless, because the limiting distribution is case dependent. Only for a known data distribution would it be possible to approach the obtained asymptotic distribution by Monte Carlo simulation.

Finally, considering part (i) of Theorem 3, we derive the asymptotic distribution of the jump size M-estimator in the next corollary.

COROLLARY 1. *Under A1.1, A1.2, and A2.1–A2.8, it holds that*

$$\sqrt{n}(\hat{\lambda}_n - \lambda) \xrightarrow{d} (\tau_0(1-\tau_0))^{-1/2} M^{-1} S^{1/2} Z_p.$$

The results corresponding to the scale invariant version of the parameter estimators, described in Remark 3, are summarized in Corollary 2, which follows. We also make the following assumptions.

A3.

(i) $\sup_t |E[\sup_{\sigma>0} \psi(\sigma^{-1}U_t)|\mathcal{F}_t^x]| = 0$, where $\mathcal{F}_t^x = \sigma(X_{-\infty},\dots,X_t)$.
(ii) $\sup_t \|X_t\|_2 < \infty$.
(iii) $\lim_{n\to\infty} n^{-1}\sum_{t=1}^{n-1}\sum_{m=1}^{n-t}\|X_t X_{t+m}\|_1 < \infty$.

COROLLARY 2. *Under Assumptions A1.1, A1.2, A2.1–A2.8, and A3 the results of Theorem 3 apply to the scale invariant estimator considered in Remark 3.*

Note that Assumption A3(i) is quite restrictive. For instance, if we consider both regressors and error i.i.d. and mutually independent, this condition will remain true if and only if we assume symmetry of the error distribution, taking into account Assumptions A1. It would be possible to develop asymptotic results along the lines of Huber (1982), which hold whether or not the errors are symmetric, but this would yield a rather unwieldy asymptotic distribution and is beyond the scope of this paper. Sufficient conditions for A3(iii) to hold can be found in Davidson (1994, p. 482).[5]

The limiting distribution in Theorem 3 has been obtained under general conditions of weak dependence for the data. However, two specific cases are worth mentioning. The first case concerns i.i.d. observations, and it could be considered a particular case of the previous one. The second one allows for trending regressors and weakly dependent perturbations. Both cases are outlined subsequently, by Assumptions A4 and A5, respectively.

A4. Let $\{(X_t', U_t)'\}_{t\le n}$ be a sequence of i.i.d. random vectors, mutually independent, and $D = E[X_t X_t']$ a nonsingular matrix.

A5. Let $\{X_t = g(t/n)\}_{t\le n}$, where g is a bounded and continuously differentiable vector valued function on $[0,1)$. The sequence $\{\psi(U_t)\}_{t\le n}$ satisfies the same requirements of $\{\eta_t\}_{t\le n}$ in A2.4 and it holds that

A5.1. Uniformly in $(\theta,\tau) \in \Theta \times \Pi$, $\lim_{n\to\infty} n^{-1}\sum_{t=1}^{[n\tau]} E[\dot\psi(U_t + \theta' X_t) g(t/n) \times g(t/n)'] = M(\theta,\tau)$, a positive definite matrix $\forall(\theta,\tau) \in \Theta \times \Pi$. Let $m(\psi) = \lim_{n\to\infty} n^{-1}\sum_{t=1}^{n} E[\dot\psi(U_t)]$.
A5.2. Uniformly in $\tau \in (0,1)$, we assume that $\lim_{n\to\infty} \mathrm{var}[n^{-1/2}\sum_{t=1}^{[n\tau]}\psi(U_t)] = \tau s(\psi)$, where $s(\psi) = \lim_{n\to\infty} \mathrm{var}[n^{-1/2}\sum_{t=1}^{n}\psi(U_t)]$, finite and positive.

Proposition 3, which follows, provides a functional central limit theorem for the shift point estimator under these two new sets of conditions. Equivalent results can also be obtained for the estimated regression parameters, but they will not presented here to save space. Interested readers are referred to Fiteni (1998).

PROPOSITION 3. *Under A1.1, A1.2, A2.1, A2.2, A2.6, and* $\lambda_n \to 0$ *with* $n\|\lambda_n\|^2 \to \infty$, *we have that*

3.1. Assuming A4, $(\lambda_n' D\lambda_n)n(\hat{\tau}_n - \tau_0) \Rightarrow c_\psi \arg\max_v\{W(v) - \frac{1}{2}|v|\}$, *with* $c_\psi = E[\psi^2(U_t)]E[\dot{\psi}(U_t)]^{-2}$.

3.2. Assuming A5, $(\lambda_n' g(\tau_0)g(\tau_0)'\lambda_n)n(\hat{\tau}_n - \tau_0) \Rightarrow \tilde{c}_\psi \arg\max_v\{W(v) - \frac{1}{2}|v|\}$, *with* $\tilde{c}_\psi = s(\psi)m(\psi)^{-2}$.

The results in this section allow us to make inferences about the model parameters. But we need to obtain consistent estimates of two matrices, M, defined in A2.7, and S, defined in A2.8. Regarding the first one, we propose to use $(\hat{M}_1 + \hat{M}_2)$, such that $\hat{M}_1 = n^{-1}\sum_1^{[n\hat{\tau}]}\dot{\psi}(Y_t - \hat{\beta}_{1n}'X_t)X_tX_t'$ and $\hat{M}_2 = n^{-1}\sum_{[n\hat{\tau}]+1}^{n}\dot{\psi}(Y_t - \hat{\beta}_{2n}'X_t)X_tX_t'$. When serial correlation is present, the second one can be consistently estimated by $(\hat{S}_1 + \hat{S}_2)$, using a kernel-based method for the partial covariance estimators $\hat{S}_j$, for $j = 1$ (with the data from 1 to $[n\hat{\tau}]$) and $j = 2$ (with the data from $[n\hat{\tau}]$ to n). This method was proposed and discussed by Andrews (1991, 1993) in a similar context. The consistency proof of the preceding matrix estimators is detailed by Fiteni (1999).

4. MONTE CARLO EXPERIMENTS

In this section, simulations are performed to verify some theoretical properties of the change point estimator for a finite sample situation. Under different distributional scenarios for the error term, we compare the performance of LS, LAD, and Huber estimators. The first is the most efficient estimator under normality; the second is the most robust; and the third one is an intermediate solution between them, providing a compromise between efficiency and robustness. We compare the estimators in terms of bias and mean square error (MSE). We also estimate their tail probabilities, obtaining the proportion of times that the estimate is found outside the asymptotic confidence intervals at different confidence levels.

Data are generated according to the following model:

$$Y_t = 1 + X_t + \lambda_n I(t/n > \tau_0) + u_t, \qquad t = 1,\dots,n,$$

where $\tau_0 = 0.5$, $x_t \sim i.i.d.N(0,1)$ and $u_t \sim i.i.d.F(u)$, with F generated as a standard normal, double exponential, t_3, t_5, and two contaminated standard normal distributions. The latter will mimic the effects of 10% and 25% outliers by taking the respective percentages on observations from $N(0,3^2)$ and the others from $N(0,1)$. These will be denoted by $N90$, and $N75$ respectively. The error term is standardized to get a variance equal to one for all cases. For the estimation of the tail probability, we consider the asymptotic distribution obtained in Theorem 3(ii). We suppose that the size of the shift is decreasing with the sample size, at a rate such that $n\|\lambda_n\|^2 \to \infty$, considering those values of λ_n that satisfy $n^{1-\delta}\|\lambda_n\|^2 = O(1)$, for $\delta \in (0,1)$. In this simulation study, we have set $\delta = \frac{1}{2}$ and, then, $\|\lambda_n\| = O(n^{-1/4})$. Thus, for $n = 100$, 200, and 500, the considered values of λ_n are 2.1892, 2.0, and 1.79527, respectively. We have fixed

$\lambda_n = 2$ for $n = 200$, and the values corresponding to the other sample sizes have been obtaining according to this.

For each type of distribution $F(u)$, 5,000 replications are performed and LS, LAD, and Huber estimators are obtained to compare them under these different scenarios. The computed Huber estimator is scale invariant, considering the MAD as the scale estimator (in Remark 3), and the corresponding constant c equals 1.345, according to the minimax version (see Huber, 1982). The programs are written in FORTRAN90 Double Precision, and the IMSL routines were used for the random number generation. We have applied the algorithm designs proposed by Koenker and D'Orey (1987) for the LAD estimation and those proposed by Huber and Dutter (1974), Dutter (1975), and Huber (1977) for the Huber estimator. Tables 1 and 2 show the results.

As regards the point estimation performance, the results are as expected in terms of MSE. In the standard normal case, LS is the best estimator for all n.

TABLE 1. Bias and mean squared error for LS, LAD, and Huber estimators of the structural break point[a]

Point Estimation		$n = 100$		$n = 200$		$n = 500$	
Model	Estimator	Bias	MSE	Bias	MSE	Bias	MSE
	LS	0.292	2.252	0.418	0.426	0.656	0.312
$N(0,1)$	LAD	0.390	3.819	0.379	0.738	0.544	0.593
	Huber	0.180	2.395	0.418	0.453	0.583	0.356
	LS	−0.364	2.166	−0.062	0.900	−0.037	0.318
$\frac{1}{2}\exp(\lvert u\rvert)$	LAD	−0.018	0.987	−0.005	0.351	−0.173	0.138
	Huber	−0.026	2.166	−0.058	0.534	−0.273	0.198
	LS	−1.544	2.972	−0.144	1.527	0.042	0.443
t_3	LAD	−0.306	0.590	−0.084	0.263	0.229	0.089
	Huber	−0.474	0.653	0.084	0.307	0.192	0.104
	LS	−1.000	2.119	−0.189	0.954	0.214	0.311
t_5	LAD	−0.246	1.610	−0.065	0.669	0.359	0.254
	Huber	−0.136	1.283	0.114	0.527	0.296	0.237
	LS	−0.156	1.943	−0.525	0.958	−0.028	0.326
$N90$	LAD	−0.070	1.339	−0.085	0.621	−0.014	0.217
	Huber	−0.184	1.099	0.027	0.535	−0.025	0.160
	LS	0.820	2.039	−0.267	0.949	−0.005	0.032
$N75$	LAD	−0.384	0.777	−0.144	0.362	−0.067	0.012
	Huber	−0.122	1.013	−0.228	0.447	−0.012	0.014

[a]Based on 5,000 replications of the model $Y_t = 1 + X_t + \lambda_n I(t/n > \tau_0) + u_t$, with $t = 1,\dots,n$ and $n = 100$, 200, and 500, such that $\lambda_n = 2.189$, 2.0, and 1.795, respectively. Several distributional scenarios for u_t are considered: standard normal, double exponential, t_3, t_5, and two mixed standard normal distribution, $N90$ and $N75$, with 10% and 25% of a normal distribution with variance equal to 9, respectively. The values corresponding to bias and MSE must be divided by 10^3.

TABLE 2. Interval estimation for LS, LAD, and Huber estimators of the structural break point[a]

Interval Estimation		$n = 100$			$n = 200$			$n = 500$		
Model	Estimator	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$
	LS	0.158	0.091	0.036	0.130	0.074	0.018	0.121	0.065	0.016
$N(0,1)$	LAD	0.120	0.073	0.026	0.088	0.044	0.009	0.088	0.046	0.009
	Huber	0.157	0.096	0.037	0.127	0.073	0.018	0.122	0.067	0.016
	LS	0.145	0.093	0.031	0.125	0.070	0.022	0.114	0.062	0.018
$\frac{1}{2}\exp(\lvert u \rvert)$	LAD	0.068	0.034	0.007	0.056	0.028	0.004	0.067	0.030	0.006
	Huber	0.148	0.086	0.026	0.118	0.067	0.018	0.114	0.061	0.016
	LS	0.149	0.087	0.032	0.125	0.075	0.025	0.111	0.064	0.016
t_3	LAD	0.052	0.021	0.004	0.058	0.022	0.003	0.057	0.024	0.004
	Huber	0.140	0.078	0.02	0.119	0.067	0.016	0.115	0.062	0.014
	LS	0.144	0.084	0.027	0.127	0.071	0.018	0.116	0.063	0.016
t_5	LAD	0.079	0.038	0.012	0.066	0.034	0.008	0.071	0.032	0.006
	Huber	0.138	0.080	0.023	0.126	0.067	0.017	0.116	0.064	0.018
	LS	0.140	0.081	0.029	0.136	0.080	0.021	0.114	0.066	0.015
$N90$	LAD	0.076	0.039	0.009	0.078	0.034	0.006	0.070	0.034	0.005
	Huber	0.144	0.084	0.026	0.135	0.078	0.023	0.113	0.061	0.016
	LS	0.147	0.085	0.032	0.133	0.077	0.021	0.111	0.065	0.018
$N75$	LAD	0.061	0.094	0.064	0.064	0.030	0.005	0.061	0.028	0.003
	Huber	0.151	0.084	0.028	0.133	0.075	0.021	0.113	0.062	0.015

[a]Based on 5,000 replications of the model $Y_t = 1 + X_t + \lambda_n I(t/n > \tau_0) + u_t$, with $t = 1, \ldots, n$ and $n = 100$, 200, and 500, such that $\lambda_n = 2.189$, 2.0, and 1.795, respectively. Several distributional scenarios for u_t are considered: standard normal, double exponential, t_3, t_5, and two mixed standard normal distribution, $N90$ and $N75$, with 10% and 25% of a normal distribution with variance equal to 9, respectively.

Similarly, the LAD estimator presents the least MSE under the double exponential distribution. With the mixed and t-distributions, the LS estimator performs comparatively rather badly, whereas the other ones have a very close behavior. The Huber estimator performs better than the LAD estimator with the distributions t_5 and $N90$ and all the sample sizes. The counterpart is obtained with the more contaminated distributions, t_3 and $N75$, although the difference is not very meaningful.

We also report the proportion of times that the estimators fall outside the asymptotic confidence intervals constructed from the asymptotic distribution (16), where the standard errors are estimated as indicated in the last paragraph of Section 3. The asymptotic critical values c_α, are equal to 7.69, 11.035, and 19.78 for $\alpha = 0.1$, 0.05, and 0.01, respectively. Inspecting Table 2, we can observe that the LAD estimator underestimates the tail probabilities in all the distributional scenarios, and this feature remains so for the largest sample size.[6] Otherwise, LS and Huber estimators approximate the probabilities rather well, obtaining a good result for $n = 500$.

Finally, it is worth mentioning that, in this simulation study, we have assumed a decreasing shift and we have compared the estimated probabilities with respect to the asymptotic distribution of the break estimator only for this case, which, as we know, is distribution free. For illustrative purposes, it could be of interest to compare them with the asymptotic distribution under the assumption of a constant shift, as we can see in Bai (1994) for the LAD estimator. In fact, Bai concludes that the approximation for this case is better than for the first one. However, doing so implies assuming that the data generating process is perfectly known, which is not suitable in practice. We have also considered other designs, which are not reported here, allowing a change in the slope of the model, the break date to be $n/4$ and $3n/4$, and a certain degree of dependency for the error term. The results do not change in a significant way. For the last case, larger samples are needed.

NOTES

1. Although the analysis of this paper is carried out in terms of a pure structural change model, such that each component of λ is nonzero, the argument can be extended to a partial structural change model without essential difficulty. The difference between them is a matter of efficiency, given that the second model incorporates (with respect to the first one) additional null restrictions about some components of λ.

2. Otherwise, the partial coefficient estimators will be defined as the minimum values of the corresponding subsets $\{\hat\beta_{1n}(\tau)\}$ and $\{\hat\beta_{2n}(\tau)\}$, respectively.

3. $\{Y_t\}$ is an L_2-*NED* process on $\{U_t\}$ with $v_m = \sum_{j=m+1}^{\infty}|\rho_1^j| + |\rho_2^j|$ and $d_t = 2\sup_{s\le t}\|U_t\|_2$. Clearly, v_m is of size $-q_0$ if $|\rho_1^j| + |\rho_2^j| = O(j^{-1-q})$ for $q > q_0$. Then, $\{Y_{t-j}\}$ will be an L_2-NED with $v_m' = v_0 I(m \le j) + v_{m-j} I(m > j)$ of size $-q_0$ and constants $d_t' = 2d_{t-j}$. Finally, observe that

$$\|\eta_t - E[\eta_t|\mathcal{F}_{t-m}^{t+m}]\|_2 \le \sup_t \|\psi(U_t)\|_2 \|Y_{t-1} - E[Y_{t-1}|\mathcal{F}_{t-m}^{t+m}]\|_2,$$

where $\mathcal{F}_{t-m}{}^{t+m} = \sigma(U_{t-m},\ldots,U_{t+m})$.

4. In fact, this result will be obtained whenever the matrix $M(\theta)$ has a constant value for all $\theta \in \Theta$.

5. I am very grateful to D.W.K. Andrews for bringing this point to my attention with very useful comments.

6. Although not reported here, the LAD estimator was also obtained considering greater samples, e.g., $n = 1{,}000$, and the approximation of the tail probabilities was really good. The previously mentioned problems, with smaller samples, can be due to the nonparametric estimation of the density function that involves the covariance estimation of the LAD estimator.

REFERENCES

Andrews, D.W.K. (1988) Laws of large numbers for dependent non-identically distributed random variables. *Econometric Theory* 4, 458–467.

Andrews, D.W.K. (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–858.

Andrews, D.W.K. (1993) Tests for parameter instability and structural change with unknown change point. *Econometrica* 61, 821–856.

Andrews, D.W.K., I. Lee, & W. Ploberger (1994) Optimal changepoint tests for normal linear regression. *Journal of Econometrics* 70, 9–38.

Andrews, D.W.K. & W. Ploberger (1994) Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica* 62, 1383–1414.

Bai, J. (1994) Least squares estimation of a shift in linear processes. *Journal of Time Series Analysis* 15, 453–472.

Bai, J. (1995) Least absolute deviation estimation of a shift. *Econometric Theory* 11, 403–436.

Bai, J. & P. Perron (1998) Estimating and testing linear models with multiple structural changes. *Econometrica* 66, 47–78.

Bhattacharya, P.K. (1987) Maximum likelihood estimation of a change-point in the distribution of independent random variables, general multiparameter case. *Journal of Multivariate Analysis* 23, 183–208.

Bierens, H.J. (1981) Robust methods and asymptotic theory. *Lecture Notes in Economics and Mathematical System*, p. 192. Berlin: Springer-Verlag.

Billingsley, P. (1968) *Convergence of Probability Measures*. New York: Wiley.

Bloomfield, P. & W.L. Steiger (1983) Least absolute deviations: Theory, applications, and algorithms. In P. Huber & M. Rosenblatt (eds.), *Progress in Probability and Statistics*. Boston: Birkduser.

Brodsky, B.S. & B.E. Darkhovsky (1993) *Nonparametric Methods in Change-Point Problems*. Dordrecht: Kluwer Academic.

Broemeling, L.D. & H. Tsurumi (1986) *Econometric and Structural Change*. New York: Marcel Dekker.

Brown, R.L., J. Durbin, & J.M. Evans (1975) Techniques for testing the constancy of regression relationships over time (with discussion). *Journal of the Royal Statistical Society, Series B* 37, 149–192.

Carlstein, E. (1988) Nonparametric estimation of a change-point. *Annals of Statistics*, 16, 188–197.

Chu, J.S. & C.K. Wu (1993) Kernel-type estimators of jump points and values of a regression function. *Annals of Statistics* 21, 1545–1566.

Csörgo M. & L. Horváth (1988) Nonparametric methods for the change-point problem. In P.R. Krishnaiah & C.R. Rao (eds.), *Handbook of Statistics,* vol. 7, pp. 403–425. Amsterdam: North Holland.

Csörgo M. & L. Horváth (1997) *Limit Theorems in Change-Point Analysis*. Wiley Series in Probability and Statistics. New York: Wiley.

Davidson, J. (1994) Stochastic limit theory. *Advanced Texts in Econometrics*. Oxford: Oxford University Press.

Delgado, M.A. & J. Hidalgo (2000) Nonparametric inference on structural breaks. *Journal of Econometrics* 96, 113–144.

Deshayes, J. & D. Picard (1986) Off-line statistical analysis of change-point models using nonparametric and likelihood methods. In M. Basseville and A. Benveniste (eds.), *Detection of Abrupt Change in Signals and Dynamical Systems*, Lecture Notes in Control and Information Sciences, no. 77, 103–168. Berlin: Springer-Verlag.

Dümbgen, L. (1991) The asymptotic behavior of some nonparametric change-point estimators. *Annals of Statistics* 19, 1471–1495.

Dutter, R. (1975) Numerical Solutions of Robust Regression Problems: Computational Aspects, a Comparison. Technical report 7, Fachgruppe für Statistik, ETH, Zürich.

Feder, P. (1975) The log likelihood ratio in segmented regression. *Annals of Statistics* 3, 84–97.

Fiteni, I. (1998) Robust Estimation of Structural Break Points. Working paper 98–51. Universidad Carlos III de Madrid.

Fiteni, I. (1999) *Inferencia en Modelos con Cambio Estructural.* Ph.D. Thesis, Department of Statistics and Econometrics, Universidad Carlos III de Madrid.

Garcia, R. & P. Perron (1996) An analysis of the real interest rate under regime shifts. *Review of Economic Studies* 78, 111–125.

Ghysels, E., A. Guay, & A. Hall (1997) Predictive tests for structural change with unknown breakpoint. *Journal of Econometrics* 82, 209–233.

Hackl, P. (ed.) (1980) *Testing the Constancy of Regression Relationships Over Time.* Göttingen: Vandenhoeck und Ruprecht.

Hackl, P. (ed.) (1989) *Statistical Analysis and Forecasting of Economic Structural Change.* Berlin: Springer-Verlag.

Hackl, P. & A. Westlund (1989) Statistical analysis of "structural change": An annotated bibliography. *Empirical Economics* 143, 167–192.

Hackl, P. & A. Westlund eds. (1991) *Economic Structural Change: Analysis and Forecasting*, 95–119. Berlin: Springer-Verlag.

Hansen, B.E. (1991) Strong laws for dependent heterogeneous processes. *Econometric Theory* 7, 213–221.

Hinkley, D.V. (1970) Inference about the change point in a sequence of random variables. *Biometrika* 57, 1–17.

Huber, P.J. (1964) Robust estimation of a location parameter. *Annals of Mathematical Statistics* 35, 73–101.

Huber, P.J. (1977) Robust methods of estimation of regression coefficients. *Mathematische Operationsforschung Statistik. Series Statistics* 8, 41–53.

Huber, P.J. (1982) *Robust Statistics.* Wiley Series in Probability and Mathematical Statistics. New York: Wiley.

Huber, P.J. & R. Dutter (1974) Numerical solutions of robust regression problems. In G. Brushmann (ed.), *COMPSTAT, Proceedings Symposium on Computational Statistics.* Berlin: Physika-Verlag, 165–172.

Hušková, M. (1990) Some asymptotic results for robust procedures for testing the constancy of regression models. *Kybernetika* 26, 392–403.

Hušková, M. (1997) Estimators of Changes. Working paper, Charles University, Praha.

Hušková, M. & P.K. Sen (1989) Nonparametric tests for shift and change in regression at an unknown time point. In P. Hackl (ed.), *Statistical Analysis and Forecasting of Economic Structural Change*, pp. 73–87. Berlin: Springer-Verlag.

Ibramigov, I.A. (1962) Some limit theorems for stationary processes. *Theory of Probability and Its Applications* 7, 349–382.

Jurečková, J. & P.K. Sen (1993) *Robust Statistical Procedures: Asymptotic and Interrelations.* New York: Wiley.

Koenker, R.W. & V. D'Orey (1987) Computing regression quantiles. *Royal Statistical Society Ser. Applied Statistics* 8, 383–393.

Krishnaiah, P.R. & B.Q. Miao (1988) Review about estimation of change points. *Handbook of Statistics,* vol. 7. New York: Elsevier.

Liu, J.S. Wu & J.V. Zidek (1997) On segmented multivariate regressions. *Statistica Sinica* 7, 497–525.

McLeish, D.L. (1975a) A maximal inequality and dependent strong laws. *Annals of Probability* 3, 829–839.

McLeish, D.L. (1975b) Invariance principles for dependent variables. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebeite* 32, 165–178.

McLeish, D.L. (1977) On the invariance principle for nonstationary mixingales. *Annals of Probability* 5, 616–621.

Picard, D. (1985) Testing and estimating change-points in time series. *Advances in Applied Probability* 17, 841–867.

Pollard, D. (1984) *Convergence of Stochastic Processes*. New York: Springer-Verlag.

Pötscher, P.M. & I.R. Prucha (1991) Basic structure of the asymptotic theory in dynamic nonlinear econometric models, part I: Consistency and approximation concepts. *Econometric Reviews* 10, 125–216.

Rothenberg, T.J. (1971) Identification in parametric models. *Econometrica* 39, 577–591.

Stock, J.H. (1997) Unit roots, structural breaks, and trends. In R.F. Engle and D.L. McFadden (eds.), *Handbook of Econometrics,* vol. 4, pp. 2739–2841. New York: Elsevier.

Wooldridge, J.M. & H. White (1988) Some invariance principles and central limit theorems for dependent heterogeneous processes. *Econometric Theory* 4, 210–230.

Yao, Y.-C. (1987) Approximating the distribution of the change-point in a sequence in independent r.v.'s. *Annals of Statistics* 15, 1321–1328.

Yohai, V.J. (1987) High breakdown-point and high efficiency robust estimates for regression. *Annals of Statistics* 15, 642–656

APPENDIX

We shall consider the case of $\tau \le \tau_0$, without loss of generality because of symmetry. Limits are taken as n, the sample size, increases to infinity. For notational convenience, we establish that $[n\tau] = k$ and $[n\tau_0] = k_0$, and $\sum_{t=i}^{j}$ will be denoted by $\sum_i^j$. The subscript n of the estimators will be omitted.

Proof of Proposition 2.

Proof of (2.1). Applying the mean value theorem (MVT) to (7),

$$0_p' = (\beta_1(\tau) - \beta_{10})'\tau M(\tilde{\theta}_1), \tag{A.1}$$

$$0_p' = (\beta_2(\tau) - \beta_{10})'(\tau_0 - \tau)M(\bar{\theta}_1) + (\beta_2(\tau) - \beta_{20})'(1 - \tau_0)M(\bar{\theta}_2), \tag{A.2}$$

for any $\tau \in \Pi$ and large n, by Assumptions A1.1, A1.2, A2.1–A2.5, and A2.7 and Lemma 1 (Lemmas 1–6 appear at the end of this Appendix), where $\tilde{\theta}_1 = \tilde{\delta}_1(\beta_1(\tau) - \beta_{10})$ and $\bar{\theta}_i = \bar{\delta}_i(\beta_2(\tau) - \beta_{i0})$, for $i = 1,2$, such that $0 < \tilde{\delta}_1, \bar{\delta}_i < 1$. By A2.7, $M(\theta)$ has full rank for all $\theta \in \Theta$, and then (A.1) and (A.2) will be satisfied if and only if

$$\beta_1(\tau) = \beta_{10},$$

$$\beta_2(\tau) = \mathcal{M}_2^{-1}((\tau_0 - \tau)M(\bar{\theta}_1)'\beta_{10} + (1 - \tau_0)M(\bar{\theta}_2)'\beta_{20}), \tag{A.3}$$

given that $\mathcal{M}_2$ is nonsingular. By adding and subtracting $(\tau_0 - \tau)\mathcal{M}_2^{-1}M(\bar{\theta}_1)'\beta_{20}$ on the right side of (A.3), we establish the result. The case of $\tau \geq \tau_0$ is obtained by symmetry. ∎

Proof of (2.2). Applying the MVT to the objective function (2),

$$\lim_{n\to\infty} E[\mathcal{S}_n] = \lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} E[\rho(U_t)] + (\beta_2(\tau) - \beta_{10})'(\tau_0 - \tau)M(\bar{\theta}_1)(\beta_2(\tau) - \beta_{10})$$

$$+ (\beta_2(\tau) - \beta_{20})'(1 - \tau_0)M(\bar{\theta}_2)(\beta_2(\tau) - \beta_{20}),$$

by Assumptions A1.1, A1.2, A2.1–A2.5, and A2.7 and Lemma 1, with $\bar{\theta}_j$, $j = 1,2$, defined as previously. By A2.7, it holds that $\forall\tau \neq \tau_0$, the second and third terms of the preceding equation are no less than zero, and at least one of them will be strictly positive. Both terms will be equal to zero only for τ_0. Thus, we identify τ_0 as $\arg\min_{\tau\in\Pi}\lim_{n\to\infty}E[\mathcal{S}_n(\tau)]$, with the objective function concentrated in τ. By (2.1), the identification of τ_0 implies the identification of the regression parameter vector ξ_0, defined by (1). ∎

Proof of Theorem 2. This theorem follows from Propositions 4–18, stated at the end of this proof. We deal simultaneously with both fixed and shrinking shifts, which will be distinguished only when needed. We want to prove that

$$\Pr\left\{\sup_{A\cup B\cup D}(\mathcal{S}_n(\xi_0) - \mathcal{S}_n(\xi)) > 0\right\} < 3\varepsilon. \tag{A.4}$$

From the definition of $\mathcal{S}_n(\xi)$, it holds that

$$\mathcal{S}_n(\xi_0) - \mathcal{S}_n(\xi) = \frac{1}{n}\sum_{t=1}^{k_0}\rho(Y_t - \beta_{10}'X_t) + \frac{1}{n}\sum_{t=k_0+1}^{n}\rho(Y_t - \beta_{20}'X_t) \tag{A.5}$$

$$-\frac{1}{n}\sum_{t=1}^{k_0}\rho(Y_t - \beta_1'X_t) - \frac{1}{n}\sum_{t=k_0+1}^{n}\rho(Y_t - \beta_2'X_t) \tag{A.6}$$

$$+\frac{1}{n}\sum_{t=k+1}^{k_0}\rho(Y_t - \beta_1'X_t) - \frac{1}{n}\sum_{t=k+1}^{k_0}\rho(Y_t - \beta_2'X_t). \tag{A.7}$$

By A1.1 and the MVT, (A.5) + (A.6) is equal to

$$-(\beta_{10} - \beta_1)'\frac{1}{n}\sum_{t=1}^{k_0}\eta_t - \frac{1}{2}(\beta_{10} - \beta_1)'\frac{1}{n}\sum_{t=1}^{k_0}\dot{\eta}_t(\delta_1(\beta_{10} - \beta_1))(\beta_{10} - \beta_1)$$

$$-(\beta_{20} - \beta_2)'\frac{1}{n}\sum_{t=k_0+1}^{n}\eta_t - \frac{1}{2}(\beta_{20} - \beta_2)'\frac{1}{n}\sum_{t=k_0+1}^{n}\dot{\eta}_t(\delta_2(\beta_{20} - \beta_2))(\beta_{20} - \beta_2),$$

such that $0 < \delta_1, \delta_2 < 1$ and $\dot{\eta}_t(\theta) = \partial \eta_t(\theta)/\partial \theta'$. Next, by adding and subtracting $n^{-1} \sum_{k+1}^{k_0} \rho(u_t)$, we obtain that (A.7) is equal to

$$(\beta_{10} - \beta_1)' \frac{1}{n} \sum_{t=k+1}^{k_0} \eta_t + \frac{1}{2} (\beta_{10} - \beta_1)' \frac{1}{n} \sum_{t=k+1}^{k_0} \dot{\eta}_t(\delta_3(\beta_{10} - \beta_1))(\beta_{10} - \beta_1)$$

$$- (\beta_{10} - \beta_2)' \frac{1}{n} \sum_{t=k+1}^{k_0} \eta_t - \frac{1}{2} (\beta_{10} - \beta_2)' \frac{1}{n} \sum_{t=k+1}^{k_0} \dot{\eta}_t(\delta_4(\beta_{10} - \beta_2))(\beta_{10} - \beta_2), \tag{A.8}$$

by the MVT, where $0 < \delta_3, \delta_4 < 1$. Next, define $M_n(j, l, \theta) = n^{-1} \sum_{j+1}^{l} \dot{\eta}_t(\theta)$ and $N_n(j, l) = n^{-1} \sum_{j+1}^{l} \eta_t$. Gathering (A.5)–(A.7) and arranging terms, we have that $\mathcal{S}_n(\xi_0) - \mathcal{S}_n(\xi)$ is equal to

$$-\frac{1}{2} (\beta_{10} - \beta_1)' M_n(0, k, \delta_1(\beta_{10} - \beta_1))(\beta_{10} - \beta_1) - (\beta_{10} - \beta_1)' N_n(0, k) \tag{A.9}$$

$$-\frac{1}{2} (\beta_{20} - \beta_2)' M_n(k_0, n, \delta_2(\beta_{20} - \beta_2))(\beta_{20} - \beta_2) - (\beta_{20} - \beta_2)' N_n(k_0, n) \tag{A.10}$$

$$-2\left(\frac{1}{2} \lambda' + \frac{1}{2} (\beta_{20} - \beta_2)'\right) M_n(k, k_0, \delta_4(\beta_{10} - \beta_2)) \left(\frac{1}{2} \lambda + \frac{1}{2} (\beta_{20} - \beta_2)\right) \tag{A.11}$$

$$-(\lambda' + (\beta_{20} - \beta_2)') N_n(k, k_0), \tag{A.12}$$

taking into account that $(\beta_{10} - \beta_2) = \lambda + (\beta_{20} - \beta_2)$. Finally, observe that the parameter space

$$\begin{aligned} A \cup B \cup D &= [(A \cup B) \cap D] \cup [(A \cup B) \cap \bar{D}] \cup [(\overline{A \cup B}) \cap D] \\ &= G_1 \cup G_2 \cup G_3, \end{aligned} \tag{A.13}$$

a union of disjoint sets, where $\bar{E}$ denotes the complementary set of E. Then, the left side of (A.4) is upper bounded by

$$\Pr\left\{\sup_{G_1} (\mathcal{S}_n(\xi_0) - \mathcal{S}_n(\xi)) > 0\right\} + \Pr\left\{\sup_{G_2} (\mathcal{S}_n(\xi_0) - \mathcal{S}_n(\xi)) > 0\right\}$$

$$+ \Pr\left\{\sup_{G_3} (\mathcal{S}_n(\xi_0) - \mathcal{S}_n(\xi)) > 0\right\},$$

with $(\mathcal{S}_n(\xi_0) - \mathcal{S}_n(\xi))$ defined by (A.9)–(A.12). Thus, it suffices to show that each of the terms composing the preceding expression is bounded by ε for both C and n large enough. This is obtained subsequently.

First Case ($G_1 = [(A \cup B) \cap D]$). To bound the probability (A.4) restricted to this parametric set, we need to consider the following possible cases of β_2:

$$(\beta_{20} - \beta_2) \geq -\lambda : \begin{cases} \textbf{(b.1)}\ F_1 = \{(\beta_{20} - \beta_2) \geq -\lambda, \|\beta_{20} - \beta_2\| > C/\sqrt{n}\}. \\ \textbf{(b.2)}\ F_2 = \{(\beta_{20} - \beta_2) \geq -\lambda, \|\beta_{20} - \beta_2\| < C/\sqrt{n}\}. \end{cases}$$

$$(\beta_{20} - \beta_2) \leq -\lambda : \begin{cases} \textbf{(b.3)}\ F_3 = \{(\beta_{20} - \beta_2) \leq -\lambda, \|\beta_{20} - \beta_2\| > C/\sqrt{n}\}. \\ \textbf{(b.4)}\ F_4 = \{(\beta_{20} - \beta_2) \leq -\lambda, \|\beta_{20} - \beta_2\| < C/\sqrt{n}\}. \end{cases}$$

Observe that $(A \cup B)$ implies $\|\beta_1 - \beta_{10}\| > C/\sqrt{n}$ and/or $\|\beta_2 - \beta_{20}\| > C/\sqrt{n}$. First, consider $\|\beta_1 - \beta_{10}\| < C/\sqrt{n}$; then we have $B \cap D$, because of G_1. In this parametric subset, the probability (A.4) will be upper bounded by

$$\Pr\left\{\sup_{\bar{A} \cap B} [(\mathrm{A.9}) + (\mathrm{A.10})] > 0\right\} + \Pr\left\{\sup_{B \cap D} [(\mathrm{A.11}) + (\mathrm{A.12})] > 0\right\},$$

where the first term is asymptotically negligible by Proposition 10. Because $B = F_1 \cup F_3$, a union of disjoint sets, an upper bound for the second term is given by

$$\Pr\left\{\sup_{F_1 \cap D} [(\mathrm{A.11}) + (\mathrm{A.12})] > 0\right\} + \Pr\left\{\sup_{F_3 \cap D} [(\mathrm{A.11}) + (\mathrm{A.12})] > 0\right\}, \qquad \textbf{(A.14)}$$

which tends to zero for large enough C, by Propositions 11 and 12. Next, we assume that $\|\beta_1 - \beta_{10}\| > C/\sqrt{n}$, so that it may happen that $B \cap D$ or $\bar{B} \cap D$. First, consider $A \cap B \cap D$, a subset of G_1 for which the probability (A.4) will be less than or equal to

$$\Pr\left\{\sup_{A} (\mathrm{A.9}) > 0\right\} + \Pr\left\{\sup_{B} (\mathrm{A.10}) > 0\right\} + \Pr\left\{\sup_{B \cap D} [(\mathrm{A.11}) + (\mathrm{A.12})] > 0\right\}.$$

First and second terms converge to zero for large enough C by Propositions 7 and 8, respectively, similar to the third one, bounded by (A.14) as in the previous case. Finally, it remains to consider the parametric subset $A \cap \bar{B} \cap D$. Observe that $\bar{B} = F_2 \cup F_4$, where $F_2 \cap F_4 = \emptyset$, and then (A.4) will be bounded by

$$\Pr\left\{\sup_{A \cap \bar{B}} [(\mathrm{A.9}) + (\mathrm{A.10})] > 0\right\} + \Pr\left\{\sup_{F_2 \cap D} [(\mathrm{A.11}) + (\mathrm{A.12})] > 0\right\}$$
$$+ \Pr\left\{\sup_{F_4 \cap D} [(\mathrm{A.11}) + (\mathrm{A.12})] > 0\right\},$$

which converges to zero by Propositions 9, 13, and 14.

Second Case ($G_2 = [(A \cup B) \cap \bar{D}]$). Proceeding as before, if $\bar{A}$ holds, then we have $B \cap \bar{D}$ because of G_2. In this case we obtain that the probability (A.4) converges to zero by Proposition 16. Next, if A occurs then we have $B \cap \bar{D}$ or $\bar{B} \cap \bar{D}$. In the first case, (A.4) is bounded by

$$\Pr\left\{\sup_{A} (\mathrm{A.9}) > 0\right\} + \Pr\left\{\sup_{B \cap \bar{D}} [(\mathrm{A.10}) + (\mathrm{A.11}) + (\mathrm{A.12})] > 0\right\},$$

and the involved terms converge to zero by Propositions 7 and 15. In the second case, i.e., $A \cap \bar{B} \cap \bar{D}$, (A.4) will be asymptotically negligible by Proposition 17.

Third Case ($G_3 = [\bar{A} \cap \bar{B} \cap D]$). In this parametric subset, the probability (A.4) converges to zero by Proposition 18.

Theorem 2 follows from Propositions 4–18, which appear subsequently. They are obtained under Assumptions A1.1, A1.2, and A2.1–A2.8. To simplify the exposition, and taking into account (A.9)–(A.12), we denote

$$\begin{aligned} M_{1n} &= M_n(0,k,\theta_1), & N_{1n} &= N_n(0,k), \\ M_{2n} &= M_n(k_0,n,\theta_2), & N_{2n} &= N_n(k_0,n), \\ M_{4n} &= M_n(k,k_0,\theta_4), & N_{4n} &= N_n(k,k_0). \end{aligned} \tag{A.15}$$

where

$$\theta_1 = \delta_1(\beta_{10} - \beta_1) = \delta_3(\beta_{10} - \beta_1), \qquad \theta_2 = \delta_2(\beta_{20} - \beta_2),$$

and $\theta_4 = \delta_4(\beta_{10} - \beta_2)$. **(A.16)**

PROPOSITION 4. *Uniformly in $\tau \in (0,1)$, it holds that*

$$M_{1n} = \tau M(\theta_1) + op(1), \qquad M_{2n} = (1-\tau_0)M(\theta_2) + op(1), \qquad \textit{and}$$

$$M_{4n} = (\tau_0 - \tau)M(\theta_4) + op(1),$$

where M_{in} and θ_i, for $i = 1$, 2, and 4, are given by (A.15) and (A.16) and $M(\cdot)$ is defined by A2.7.

Proof. Under A2.1,

$$\sup_{\tau \in \Pi} \|M_{1n} - \tau M(\theta_1)\| \leq \sup_{\tau \in \Pi} \sup_{\theta \in \Theta} \left\| \frac{1}{n} \sum_{t=1}^{[n\tau]} \dot{\eta}_t(\theta) - \tau M(\theta) \right\| \to 0,$$

by Lemma 1. The remaining results are established similarly. ■

PROPOSITION 5. *$\forall \alpha \in \mathbb{R}^p$ and $\tau \in (0,1)$, it holds that*

$$\sqrt{n}\alpha' N_{1n} \Rightarrow (\alpha' S \alpha)^{1/2} B(\tau), \qquad \sqrt{n}\alpha' N_{2n} \Rightarrow (\alpha' S \alpha)^{1/2} B(1-\tau_0), \qquad \textit{and}$$

$$\sqrt{n}\alpha' N_{4n} \Rightarrow (\alpha' S \alpha)^{1/2} B(\tau_0 - \tau),$$

where N_{in} is defined in (A.15) for $i = 1$, 2, and 4 and $B(\cdot)$ is a p-vector of independent Brownian motion processes in $[0,1]$.

Proof. Lemma 4 and the continuous mapping theorem yield the result. ■

PROPOSITION 6. *Let $M(\theta)$ be a matrix defined $\forall \theta \in \Theta$ by A2.7, with eigenvalues $a_j(\theta)$, $j = 1,\ldots,p$. Then $\forall \gamma \in \mathbb{R}^p$, it holds that $\gamma' M(\theta)\gamma \geq q(\theta)\|\gamma\|^2 > 0$, where $q(\theta) = \min_j\{a_j(\theta)\}$.*

Proof. By A2.7, $M(\theta)$ is a $p \times p$ finite and positive definite matrix for each $\theta \in \Theta$. Hence, the Courant theorem yields

$$\gamma' M(\theta)\gamma \geq \inf_{\gamma} \left\{ \frac{\gamma' M(\theta)\gamma}{\gamma'\gamma} \right\} \gamma'\gamma = \min_j \{a_j(\theta)\} \|\gamma\|^2 > 0.$$ ■

PROPOSITION 7. *Given the event* $\{E_1\} = \left\{\sup_A (A.9) > 0\right\}$*, it holds that, for large* n*,* $\Pr\{E_1\} < \varepsilon$ *for* C *large enough.*

Proof. Observe that, for large n, $\{E_1\}$ implies the event

$$\left\{\sup_A \left|\frac{(\beta_{10}-\beta_1)'\sqrt{n}N_{1n}}{\|\beta_{10}-\beta_1\|}\right| > \inf_A \frac{\sqrt{n}\|(\tau_1 M(\theta_1))^{1/2}(\beta_{10}-\beta_1)\|^2}{\|\beta_{10}-\beta_1\|}\right\}, \tag{A.17}$$

from Proposition 4 and A2.7, where $\tau_1 = \inf\{\tau : \tau \in \Pi\} > 0$. Proposition 6 yields

$$\left\{\sup_A \left|\frac{(\beta_{10}-\beta_1)'\sqrt{n}N_{1n}}{\|\beta_{10}-\beta_1\|}\right| > \tau_1 q_1 \sqrt{n}\inf_A \|\beta_{10}-\beta_1\|\right\},$$

where q_1 is the smallest eigenvalue of $M(\theta_1)$. Hence, we obtain that $\Pr\{E_1\}$ will be upper bounded by

$$\Pr\left\{\sup_A \left|\frac{(\beta_{10}-\beta_1)'\sqrt{n}N_{1n}}{\|\beta_{10}-\beta_1\|}\right| > \tau_1 q_1 C\right\} = O\left(\frac{1}{C^2}\right),$$

for large n, by Lemma 5. From here the conclusion is standard because C can be arbitrarily large. ■

PROPOSITION 8. *Given the event* $\{E_2\} = \left\{\sup_B (A.10) > 0\right\}$*, it holds that, for large* n*,* $\Pr\{E_2\} < \varepsilon$ *for* C *large enough.*

Proof. By Propositions 4 and 5 and Lemma 6, the result follows as in Proposition 7. ■

PROPOSITION 9. *Given the event* $\{E_3\} = \left\{\sup_{A\cap\bar{B}} [(A.9) + (A.10)] > 0\right\}$*, it holds that, for large* n*,* $\Pr\{E_3\} < \varepsilon$ *for* C *large enough.*

Proof. Because in the subset $\bar{B}$, $(\beta_{20}-\beta_2) = Op(n^{-1/2})$, we have $M_{2n} = Op(1)$ and $\sqrt{n}N_{2n} = Op(1)$ by Propositions 4 and 5, respectively. Then, we prove the result as in Proposition 7. ■

PROPOSITION 10. *Given the event* $\{E_4\} = \left\{\sup_{\bar{A}\cap B} [(A.9) + (A.10)] > 0\right\}$*, it holds that, for large* n*,* $\Pr\{E_4\} < \varepsilon$ *for large enough* C*.*

Proof. Using Propositions 4 and 5, the result follows as in Proposition 9. ■

PROPOSITION 11. *Given the event* $\{E_6\} = \left\{\sup_{F_1\cap D} [(A.11) + (A.12)] > 0\right\}$*, it holds that, for large* n*,* $\Pr\{E_6\} < \varepsilon$ *for* C *large enough.*

Proof. Let q_{4n} be the least eigenvalue of $nM_{4n}/(k_0 - k)$, a finite and positive definite matrix by A2.6. Then, $\{E_6\}$ will imply that

$$\left\{\sup_{F_1\cap D}\left|\frac{\lambda' + (\beta_{20}-\beta_2)'}{\|\lambda + (\beta_{20}-\beta_2)\|}\frac{nN_{4n}}{(k_0-k)}\right| > \inf_{F_1\cap D}\left|\frac{q_{4n}}{2}\frac{\|\lambda + (\beta_{20}-\beta_2)\|^2}{\|\lambda\| + \|\beta_{20}-\beta_2\|}\right|\right\}. \tag{A.18}$$

In F_1, $(\beta_{20} - \beta_2) \geq -\lambda$, and given $\lambda > 0$, it holds that

$$\|\lambda + (\beta_{20} - \beta_2)\|^2 > \|\beta_{20} - \beta_2\|^2 - \|\lambda\|^2 = (\|\beta_{20} - \beta_2\| + \|\lambda\|)(\|\beta_{20} - \beta_2\| - \|\lambda\|). \tag{A.19}$$

Hence, by Lemma 3, there exists $K < \infty$ such that

$$\Pr\{E_6\} \leq K\left(\frac{Cq_{4n}{}^2}{4\|\lambda\|^2}\left(\frac{C^{1/2}}{\sqrt{n}} - \|\lambda\|\right)^2\right)^{-1} = O\left(\frac{1}{C}\right),$$

for large n, an $o(1)$ term for arbitrarily large C. ■

PROPOSITION 12. *Given the event* $\{E_7\} = \left\{\sup_{F_3 \cap D} [(\text{A.11}) + (\text{A.12})] > 0\right\}$, *it holds that, for large n,* $\Pr\{E_7\} < \varepsilon$ *for C large enough.*

Proof. In F_3, $(\beta_{20} - \beta_2) \leq -\lambda$ and noting that $(\beta_{20} - \beta_2) < 0$, we have

$$\|\lambda + (\beta_{20} - \beta_2)\|^2 > \|\lambda\|^2 - \|\beta_{20} - \beta_2\|^2 = (\|\lambda\| + \|\beta_{20} - \beta_2\|)(\|\lambda\| - \|\beta_{20} - \beta_2\|). \tag{A.20}$$

Using similar arguments as in proving Proposition 11, we obtain that, by Lemma 3, there exists a constant $K < \infty$ such that

$$\Pr\{E_7\} \leq K\left(\frac{Cq_{4n}{}^2}{4\|\lambda\|^2}\left(\|\lambda\| - \frac{C^{1/2}}{\sqrt{n}}\right)^2\right)^{-1} = O\left(\frac{1}{C}\right),$$

for large n. The result follows from standard arguments. ■

PROPOSITION 13. *Given the event* $\{E_8\} = \left\{\sup_{F_2 \cap D} [(\text{A.11}) + (\text{A.12})] > 0\right\}$, *it holds that, for large n,* $\Pr\{E_8\} < \varepsilon$ *for C large enough.*

Proof. As in Proposition 11, given F_2, the event $\{E_8\}$ will imply

$$\left\{\sup_{F_2 \cap D}\left|\frac{(\lambda' + (\beta_{20} - \beta_2)')nN_{4n}}{\|\lambda\|(k_0 - k)}\right| > \inf_{F_2 \cap D} \frac{q_{4n}\|\lambda + (\beta_{20} - \beta_2)\|^2}{2\|\lambda\|}\right\},$$

by A2.6 and using (A.19). In the subset D, from Lemma 3, it follows that, for a constant $K < \infty$, $\Pr\{E_8\}$ will be upper bounded by

$$K\left(\frac{Cq_{4n}^2}{4\|\lambda\|^2}\left(\frac{Op(n^{-1}) - \|\lambda\|^2}{\|\lambda\|}\right)^2\right)^{-1} = K\left(\frac{Cq_{4n}^2}{4}(Op(n^{-1}\|\lambda\|^{-2}) - 1)^2\right)^{-1} = O\left(\frac{1}{C}\right),$$

for large n. Standard arguments imply the result. ■

PROPOSITION 14. *Given the event* $\{E_9\} = \left\{ \sup_{F_4 \cap D} [(\text{A.11}) + (\text{A.12})] > 0 \right\}$, *it holds that, for large n,* $\Pr\{E_9\} < \varepsilon$ *for C large enough.*

Proof. The proof follows by A2.6, Lemma 3, and (A.20) using similar arguments to those of Proposition 13. ■

PROPOSITION 15. *Given the event* $\{E_{10}\} = \left\{ \sup_{B \cap \bar{D}} [(\text{A.10}) + (\text{A.11}) + (\text{A.12})] > 0 \right\}$, *it holds that, for large n,* $\Pr\{E_{10}\} < \varepsilon$ *for large enough C.*

Proof. Observe that $\{E_{10}\}$ implies the event

$$\left\{ \sup_{B \cap \bar{D}} \left[\frac{(\beta_{20} - \beta_2)' \sqrt{n} N_{2n}}{\|\beta_{20} - \beta_2\|} + \frac{\lambda' + (\beta_{20} - \beta_2)' \sqrt{n} N_{4n}}{\|\beta_{20} - \beta_2\|} \right.\right.$$
$$+ \frac{\sqrt{n}(\beta_{20} - \beta_2)' M_{2n} (\beta_{20} - \beta_2)}{2\|\beta_{20} - \beta_2\|}$$
$$\left.\left. + \frac{2\sqrt{n}\left(\frac{1}{2}\lambda' + \frac{1}{2}(\beta_{20} - \beta_2)'\right) M_{4n} \left(\frac{1}{2}\lambda + \frac{1}{2}(\beta_{20} - \beta_2)\right)}{\|\beta_{20} - \beta_2\|} \right] > 0 \right\}. \quad \textbf{(A.21)}$$

In the subset $\bar{D}$, it holds that $|\tau - \tau_0| < C/(n\|\lambda\|^2)$, and therefore, $\forall \alpha \in \mathbb{R}^p$, $\alpha' M_{4n} \alpha = Op(n^{-1}\|\lambda\|^{-2}\|\alpha\|^2)$, and $\sqrt{n}\alpha' N_{4n} = Op(n^{-1/2}\|\lambda\|^{-1}\|\alpha\|)$, by Propositions 4 and 5 and Markov's inequality. Thus, by A2.6, A2.7, and Proposition 6, (A.21) will imply the event

$$\left\{ \sup_B \left| \frac{(\beta_{20} - \beta_2)'}{\|\beta_{20} - \beta_2\|} \sqrt{n} N_{2n} + op(1) \right| > \inf_B \left| \sqrt{n} \|\beta_{20} - \beta_2\| \left((1 - \tau_0) q_2 + Op\left(\frac{1}{C^2} + \frac{1}{n\|\lambda\|^2} \right) \right) \right| \right\},$$

and then, for large n, we obtain that $\Pr\{E_{10}\}$ will be less than or equal to

$$\Pr\left\{ \sup_B \left| \frac{(\beta_{20} - \beta_2)'}{\|\beta_{20} - \beta_2\|} \sqrt{n} N_{2n} \right| > C\left((1 - \tau_0) q_2 + Op\left(\frac{1}{C^2} \right) \right) \right\} = O\left(\frac{1}{C^2} \right),$$

from Lemma 5. Standard arguments conclude the proof. ■

PROPOSITION 16. *Given the event* $\{E_{11}\} = \left\{ \sup_{\bar{A} \cap B \cap \bar{D}} [(\text{A.9}) + (\text{A.10}) + (\text{A.11}) + (\text{A.12})] > 0 \right\}$, *it holds that, for large n,* $\Pr\{E_{11}\} < \varepsilon$ *for large enough C.*

Proof. Given $\bar{A}$, $(\beta_{10} - \beta_1) = Op(n^{-1/2})$ and then $M_{1n} = Op(1)$ and $\sqrt{n} N_{1n} = Op(1)$ by Propositions 4 and 5, respectively. Hence, the preceding event will imply that

$$\left\{ \sup_{B\cap\bar{D}} \left[\frac{(\beta_{20}-\beta_2)'}{\|\beta_{20}-\beta_2\|}\sqrt{n}N_{2n} + \frac{\lambda' + (\beta_{20}-\beta_2)'}{\|\beta_{20}-\beta_2\|}\sqrt{n}N_{4n} \right.\right.$$

$$+ \frac{\sqrt{n}(\beta_{20}-\beta_2)'M_{2n}(\beta_{20}-\beta_2)}{2\|\beta_{20}-\beta_2\|}$$

$$+ \frac{2\sqrt{n}\left(\frac{1}{2}\lambda' + \frac{1}{2}(\beta_{20}-\beta_2)'\right)M_{4n}\left(\frac{1}{2}\lambda + \frac{1}{2}(\beta_{20}-\beta_2)\right)}{\|\beta_{20}-\beta_2\|}$$

$$\left.\left. + \frac{Op(n^{-1/2})}{\|\beta_{20}-\beta_2\|} \right] > 0 \right\}.$$

Because $Op(n^{-1/2})/\|\beta_{20} - \beta_2\| = Op(1/C)$ in the subset B, the result follows as in Proposition 15. ■

PROPOSITION 17. *Given the event* $\{E_{12}\} = \left\{ \sup_{A\cap\bar{B}\cap\bar{D}} [(A.9) + (A.10) + (A.11) + (A.12)] > 0 \right\}$, *it holds that, for large n,* $\Pr\{E_{12}\} < \varepsilon$ *for large enough C.*

Proof. By Propositions 4 and 5, the result follows as in Proposition 16. ■

PROPOSITION 18. *Given the event* $\{E_{13}\} = \left\{ \sup_{\bar{A}\cap\bar{B}\cap D} [(A.9) + (A.10) + (A.11) + (A.12)] > 0 \right\}$, *it holds that, for large n,* $\Pr\{E_{12}\} < \varepsilon$ *for large enough C.*

Proof. Proposition 5 yields $\sup_{\bar{A}\cap\bar{B}\cap D}(\|\lambda\|(k_0 - k))^{-1}(\beta_{j0} - \beta_j)'nN_{jn} = Op(C^{-1})$, for $j = 1,2$ and $\{E_{13}\}$ this will imply that

$$\left\{ \sup_{\bar{B}\cap D} \left[\frac{(\lambda' + (\beta_{20}-\beta_2)')nN_{4n}}{\|\lambda\|(k_0-k)} + \frac{2\left(\frac{1}{2}\lambda' + (\beta_{20}-\beta_2)'\right)nM_{4n}\left(\frac{1}{2}\lambda + \frac{1}{2}(\beta_{20}-\beta_2)\right)}{\|\lambda\|(k_0-k)} \right.\right.$$

$$\left.\left. + \frac{n\sup_{j=1,2}\{(\beta_{j0}-\beta_j)'(M_{1n}+M_{2n})(\beta_{j0}-\beta_j)\}}{2\|\lambda\|(k_0-k)} + Op\left(\frac{1}{C}\right) > 0 \right\}.$$

From Proposition 4, $(M_{1n} + M_{2n})$ converges in probability, uniformly in τ, to $\tau M(\theta_1) + (1 - \tau_0)M(\theta_2) \geq \inf_{\theta\in\Theta} M(\theta) + (\tau - \tau_0)\inf_{\theta\in\Theta} M(\theta)$. Then, for large n, the preceding event implies that

$$\left\{ \sup_{D} \left| \frac{(\lambda' + (\beta_{20}-\beta_2)')nN_{4n}}{\|\lambda\|(k_0-k)} + Op\left(\frac{1}{C}\right) \right| \right.$$

$$\left. > \inf_{\bar{A}\cap\bar{B}\cap D} \left| \frac{q_{4n}\|\lambda + (\beta_{20}-\beta_2)\|^2}{2\|\lambda\|(k_0-k)} + \frac{(q_{\inf} + op(1))Op(n^{-1})}{2\|\lambda\|} \right| \right\},$$

by A2.7, where $q_{\inf} > 0$ is the smallest eigenvalue of $\inf_{\theta\in\Theta} M(\theta)$. Given that $\bar{B} = F_2 \cup F_4$, the result follows much as in Propositions 13 and 14 for large n and C. ■

Proof of Theorem 3. Again, we only consider the case of $v_3 < 0$, without loss of generality because of symmetry. For notational convenience, $[v_3 P_\lambda]$ will be denoted by $v_3 P_\lambda$. By (A.9)–(A.12), we obtain that

$$n\Lambda_n(v) = -\frac{1}{2}v_1' \frac{1}{n}\sum_{t=1}^{k_0} \dot{\eta}_t\left(\delta_1 \frac{v_1}{\sqrt{n}}\right)v_1 - v_1' \frac{1}{\sqrt{n}}\sum_{t=1}^{k_0} \eta_t \tag{A.22}$$

$$+ \frac{1}{2}v_1' \frac{1}{n}\sum_{t=k_0+v_3P_\lambda+1}^{k_0} \dot{\eta}_t\left(\delta_1 \frac{v_1}{\sqrt{n}}\right)v_1 + v_1' \frac{1}{\sqrt{n}}\sum_{t=k_0+v_3P_\lambda+1}^{k_0} \eta_t \tag{A.23}$$

$$- \frac{1}{2}v_2' \frac{1}{n}\sum_{t=k_0+1}^{n} \dot{\eta}_t\left(\delta_2 \frac{v_2}{\sqrt{n}}\right)v_2 - v_2' \frac{1}{\sqrt{n}}\sum_{t=k_0+1}^{n} \eta_t$$

$$- \frac{1}{2}\lambda' \sum_{t=k_0+v_3P_\lambda+1}^{k_0} \dot{\eta}_t\left(\delta_4\left(\lambda + \frac{v_2}{\sqrt{n}}\right)\right)\lambda$$

$$- \frac{1}{2}v_2' \frac{1}{n}\sum_{t=k_0+v_3P_\lambda+1}^{k_0} \dot{\eta}_t\left(\delta_4\left(\lambda + \frac{v_2}{\sqrt{n}}\right)\right)v_2$$

$$- \lambda' \frac{1}{\sqrt{n}}\sum_{t=k_0+v_3P_\lambda+1}^{k_0} \dot{\eta}_t\left(\delta_4\left(\lambda + \frac{v_2}{\sqrt{n}}\right)\right)v_2$$

$$- \lambda' \sum_{t=k_0+v_3P_\lambda+1}^{k_0} \eta_t - v_2' \frac{1}{\sqrt{n}}\sum_{t=k_0+v_3P_\lambda+1}^{k_0} \eta_t,$$

where (A.22) and (A.23) follow, noting that $\sum_1^k = \sum_1^{k_0} - \sum_{k+1}^{k_0}$. Next, using Lemma 2 and Assumption A2.1, we can rewrite

$$n\Lambda_n(v) = -\frac{1}{2}v_1'(\tau_0 M + op(1))v_1 - \frac{1}{2}v_2'((1-\tau_0)M + op(1))v_2 \tag{A.24}$$

$$+ \frac{1}{2}v_1'\left(\frac{v_3P_\lambda}{n}M + op(1)\right)v_1 - \frac{1}{2}\lambda'(v_3P_\lambda M(\lambda) + op(1))\lambda \tag{A.25}$$

$$- \frac{1}{2}v_2'\left(\frac{v_3P_\lambda}{n}M(\lambda) + op(1)\right)v_2 - \lambda' \frac{1}{\sqrt{n}}(v_3P_\lambda M(\lambda) + op(1))v_2 \tag{A.26}$$

$$- v_1' \frac{1}{\sqrt{n}}\sum_{t=1}^{k_0} \eta_t - v_2' \frac{1}{\sqrt{n}}\sum_{t=k_0+1}^{n} \eta_t + v_1' \frac{1}{\sqrt{n}}\sum_{t=k_0+v_3P_\lambda+1}^{k_0} \eta_t \tag{A.27}$$

$$- \lambda' \sum_{t=k_0+v_3P_\lambda+1}^{k_0} \eta_t - v_2' \frac{1}{\sqrt{n}}\sum_{t=k_0+v_3P_\lambda+1}^{k_0} \eta_t. \tag{A.28}$$

Because $n^{-1}P_\lambda = O(n^{-1}\|\lambda\|^{-2})$, an $o(1)$ term in both cases of λ, fixed and decreasing with n,

$$(\text{A.24}) + (\text{A.25}) + (\text{A.26}) = -\frac{1}{2}v_1'\tau_0 M v_1 - \frac{1}{2}v_2'(1-\tau_0)Mv_2 - \frac{1}{2}\lambda' M(\lambda)\lambda P_\lambda v_3 + op(1). \qquad \textbf{(A.29)}$$

Next, consider (A.27) and (A.28). Using Lemma 4, we have the following convergence results:

$$v_1'\frac{1}{\sqrt{n}}\sum_{t=1}^{k_0}\eta_t \Rightarrow v_1' S^{1/2}B(\tau_0),$$

$$v_2'\frac{1}{\sqrt{n}}\sum_{t=k_0+1}^{n}\eta_t \Rightarrow v_2' S^{1/2}B(1-\tau_0) \quad \text{and} \quad v_j'\frac{1}{\sqrt{n}}\sum_{t=k_0+v_3P_\lambda+1}^{k_0}\eta_t = op(1),$$

for $j = 1,2$. For the remaining terms, we analyze separately each case of λ, in parts (a) and (b), which follow.

(a) When $\lambda = \lambda_n$, such that $n\|\lambda_n\|^2 \to \infty$,

$$\lambda_n' \sum_{t=k_0+v_3P_{\lambda_n}+1}^{k_0}\eta_t \Rightarrow \sqrt{P_{\lambda_n}}\lambda_n' S^{1/2}W_1(-v_3) = -W_1(-P_{\lambda_n}\lambda_n' S\lambda_n v_3) = W_1(-v_3), \qquad \textbf{(A.30)}$$

using Lemma 4, for $v_3 < 0$, with a rescaling ($n^{-1/2}$ is replaced by $P_{\lambda_n}^{-1/2}$ because $k = k_0 + v_3P_{\lambda_n}$ and the number of elements in V_N is not larger than NP_{λ_n}). The last equality establishes that $P_{\lambda_n} = (\lambda_n' S\lambda_n)^{-1}$ and $W_1(\cdot)$ is a Brownian motion process on the positive half line. The counterpart of (A.30) in the case of $v_3 > 0$ has a limit $W_2(\cdot)$, another Brownian motion process defined on $\mathbb{R}^+$. The two processes are independent by the proof of Lemma 4, part (ii), because they involve nonoverlapping series. Furthermore, the third term of (A.29),

$$-\frac{1}{2}\lambda_n' v_3 P_{\lambda_n} M(\lambda_n)\lambda_n = -\frac{1}{2}\lambda_n' M\lambda_n v_3(\lambda_n' S\lambda_n)^{-1} = \frac{1}{2}\frac{\lambda_n' M\lambda_n}{\lambda_n' S\lambda_n}|v_3|,$$

by Lemma 2. The same result holds when $v_3 > 0$ by symmetry. Thus,

$$(\text{A.24}) + (\text{A.25}) + (\text{A.26}) = -\frac{1}{2}v_1'\tau_0 M v_1 - \frac{1}{2}v_2'(1-\tau_0)Mv_2 + \frac{1}{2}\frac{\lambda_n' M\lambda_n}{\lambda_n' S\lambda_n}|v_3| + op(1), \qquad \textbf{(A.31)}$$

$$(\text{A.27}) + (\text{A.28}) \Rightarrow -v_1' S^{1/2}B(\tau_0) - v_2' S^{1/2}B(1-\tau_0) - W(v_3), \qquad \textbf{(A.32)}$$

where $W(\cdot)$ is a two-sided Brownian motion defined on $\mathbb{R}$ and based on the two independent processes $W_1(\cdot)$ and $W_2(\cdot)$, previously stated.

(b) When λ is constant: Consider the process in (18) and let $W^\#(k) = \sum_{k+1}^{k_0}\eta_t$, for $k \le k_0$ (taking $W^\#(k_0) = 0$) and $W^\#(k) = \sum_{k_0+1}^{k}\eta_t$, for $k > k_0$. Thus, $W^\#(k)$ has the same distribution as $W^*(k - k_0)$, where $W^*(\cdot)$ is defined by (18). For

$k - k_0 = v_3$ and $v_3 < 0$, $-\lambda' \sum_{k_0+v_3+1}^{k_0} \eta_t \stackrel{d}{=} -\lambda' W^*(v_3)$. For $v_3 > 0$, the limit distribution is similarly defined by (18). Therefore,

$$\text{(A.24)} + \text{(A.25)} + \text{(A.26)} = -\frac{1}{2} v_1' \tau_0 M v_1 - \frac{1}{2} v_2'(1-\tau_0) M v_2 - \frac{1}{2} \lambda' M(\lambda) \lambda v_3 + op(1), \quad \textbf{(A.33)}$$

$$\text{(A.27)} + \text{(A.28)} \Rightarrow -v_1' S^{1/2} B(\tau_0) - v_2' S^{1/2} B(1-\tau_0) - \lambda' W^*(-v_3). \quad \textbf{(A.34)}$$

And the separate treatment for the cases (a) and (b) is concluded.

Considering the preceding limit distribution of the objective function, a weak convergence result for the estimators, defined in the compact set $\{|v_i| < M, i = 1,2,3\}$, can be derived using the mapping continuous theorem as follows.

Proof of (i). $\hat{v}_1$ and $\hat{v}_2$ are obtained from the first-order conditions,

$$\tau_0 M \hat{v}_1 + S^{1/2} B(\tau_0) = 0_p \quad \text{and} \quad (1-\tau_0) M \hat{v}_2 + S^{1/2} B(1-\tau_0) = 0_p. \qquad \blacksquare$$

Proof of (ii). Assuming a shrinking shift,

$$\hat{v}_3 \Rightarrow \arg\min_v \left\{ -W(v) + \frac{1}{2} \frac{\lambda_n' M \lambda_n}{\lambda_n' S \lambda_n} |v| \right\} \stackrel{d}{=} \arg\max_v \left\{ W(v) - \frac{1}{2} \frac{\lambda_n' M \lambda_n}{\lambda_n' S \lambda_n} |v| \right\}.$$

by (A.31) and (A.32). Using the change of variable $w = (\lambda_n' M \lambda_n / \lambda_n' S \lambda_n)^2 v$,

$$P_{\lambda_n}^{-1}(\hat{k} - k_0) \Rightarrow \left(\frac{\lambda_n' S \lambda_n}{\lambda_n' M \lambda_n} \right)^2 \arg\max_w \left\{ W(w) - \frac{1}{2} |w| \right\}. \qquad \blacksquare$$

Proof of (iii). For a constant λ, we consider (A.33) and (A.34), and hence,

$$\hat{v}_3 \Rightarrow \arg\min_v \left\{ -\lambda' W^*(v) + \frac{1}{2} \lambda' M(\lambda) \lambda |v| \right\} \stackrel{d}{=} \arg\max_v \left\{ \lambda' W^*(v) - \frac{1}{2} \lambda' M(\lambda) \lambda |v| \right\}.$$

$\blacksquare$

Proof of (iv). This is immediate from (A.31)–(A.34). $\blacksquare$

Proof of Corollary 1. Define the vector $H = [I_p \vdots - I_p] \in \mathbb{R}^{p \times 2p}$. Then, $\sqrt{n}(\hat{\lambda} - \lambda) = \sqrt{n} H((\hat{\beta}_1 - \beta_{10})', (\hat{\beta}_2 - \beta_{20})')'$, which converges in distribution to

$$H \begin{pmatrix} \tau_0^{-1/2} M^{-1} S^{1/2} Z_p & 0_{p\times p} \\ 0_{p\times p} & (1-\tau_0)^{-1/2} M^{-1} S^{1/2} Z_p \end{pmatrix} \stackrel{d}{=} (\tau_0(1-\tau_0))^{-1/2} M^{-1} S^{1/2} Z_p,$$

by Theorem 3. $\blacksquare$

Proof of Corollary 2. Given $\hat{s}_n \to^p \sigma$, with σ defined by A0, it suffices to show that

$$\sup_\tau \sup_\theta \left\| \frac{1}{n} \sum_{t=1}^{[n\tau]} (\dot{\eta}_t(\theta, \hat{s}_n) - E[\dot{\eta}_t(\theta, \sigma)]) \right\| \to^p 0 \quad \textbf{(A.35)}$$

and

$$\sup_{\tau}\left\|\frac{1}{\sqrt{n}}\sum_{t=1}^{[n\tau]}\eta_t(0_p,\hat{s}_n)-\frac{1}{\sqrt{n}}\sum_{t=1}^{[n\tau]}\eta_t(0_p,\sigma)\right\|\to^p 0, \tag{A.36}$$

where $\dot{\eta}_t(a,b) = \dot{\psi}(b^{-1}(U_t + a'X_t))X_tX_t'$ and $\eta_t(a,b) = \psi(b^{-1}(U_t + a'X_t))X_t$, with $a \in \mathbb{R}^p$ and $b \in \mathbb{R}$. First consider (A.35). The left-hand side is upper bounded by

$$\sup_{\tau}\sup_{\substack{\theta\\ s\in[\sigma-\varepsilon,\sigma+\varepsilon]}}\left\|\frac{1}{n}\sum_{t=1}^{[n\tau]}(\dot{\eta}_t(\theta,s)-E[\dot{\eta}_t(\theta,s)])\right\|,$$

for all $\varepsilon > 0$, which converges to zero by Lemma A3 of Andrews (1993), because $\sup_{t\le n}E[\sup_{\substack{\theta\in\Theta\\ s>0}}\|\dot{\eta}_t(\theta,s)\|^2]$ is finite by A1 and A3. Next, let X_{tj} be the jth coordinate of the regressors vector $X_t \in \mathbb{R}^p$, for $t = 1,\dots,n$, and define, for $0 \le h \le 1$, $A_{n,j}(h,\tau) = n^{-1/2}\sum_{t=1}^{[n\tau]}\psi((0{,}5\sigma + h\sigma)^{-1}U_t)X_{tj}$, which, for each $\tau \in \Pi$, belongs to C, the space of continuous functions defined in $[0,1]$. Because $\hat{s}_n \to^p \sigma$, to establish (A.36), it will be enough to show that $A_{n,j}(h,\tau)$ is tight in h uniformly in τ, for $1 \le j \le p$. From Theorem 12.3 of Billingsley (1968), we must obtain that (i) $A_{n,j}(0,\tau)$ is tight uniformly in τ, (ii) there exist $\gamma > 0$, $\delta > 0$, and a continuous and bounded function in $\mathbb{R}$ such that, for any $0 \le h_1 \le h_2 \le 1$, it holds that as $n \to \infty$,

$$P(|A_{n,j}(h_2,\tau)-A_{n,j}(h_1,\tau)|>\kappa)\le\frac{|R(h_2)-R(h_1)|^\delta}{\kappa^\gamma}, \tag{A.37}$$

uniformly in τ, for all positive κ. Part (i) is straightforwardly obtained using Lemma 4. By the Chebyshev inequality, to establish part (ii), it would be enough to obtain that, for $\gamma = 2$ and a constant $K < \infty$, $\sup_\tau E|A_{n,j}(h_2,\tau) - A_{n,j}(h_1,\tau)|^2 \le K|h_2 - h_1|^2$. According to A3(i), we have asymptotically that

$$\begin{aligned}
&\sup_{\tau} E|A_{n,j}(h_2,\tau)-A_{n,j}(h_1,\tau)|^2\\
&\quad=\sup_{\tau}\frac{1}{n}\sum_{t=1}^{[n\tau]}E\left[\left(\psi\left(\frac{U_t}{0{,}5\sigma+h_2\sigma}\right)-\psi\left(\frac{U_t}{0{,}5\sigma+h_1\sigma}\right)\right)^2X_{tj}^2\right]\\
&\quad\le\sup_{u}|\dot{\psi}(u)|^2\frac{|h_1-h_2|^2}{(0{,}5)^4\sigma^2}\sup_{\tau}\frac{1}{n}\sum_{t=1}^{[n\tau]}E[U_t^2X_{tj}^2I(U_t\le 1.5\sigma m)],
\end{aligned}$$

from MVT and A1. For $K = (6m\sup_u|\dot{\psi}(u)|\sup_t\|X_{tj}\|_2)^2$, a finite constant by A3(ii), the result follows. ■

Proof of Proposition 3. For all t, define $\psi_t = \psi(U_t)$, $\dot{\psi}_t = \dot{\psi}(U_t)$ and let $m \equiv m(\psi)$ and $s \equiv s(\psi)$.

Proof of (3.1). Under A4, $M = E[\dot{\psi}_t]D$ and $S = E[\psi_t^2]D$. Then, $(\lambda_n'S\lambda_n)^{-1} \times (\lambda_n'M\lambda_n)^2 = E[\dot{\psi}_t]^2E[\psi_t^2]^{-1}(\lambda_n'D\lambda_n)$. The result follows immediately from Theorem 3 (for which the weak convergence result in Lemma 4 can be obtained by the invariance principle for independent random variables [see, e.g., Billingsley, 1968]).

Proof of (3.2). Under A5, trending regressors $X_t = g(t/n)$ satisfy all the conditions required in Theorem 2. Hence, we concentrate on the limiting process of

$$-\lambda_n' \sum_{t=k_0+v_3P_{\lambda_n}+1}^{k_0} \eta_t - \frac{1}{2}\lambda_n' \sum_{t=k_0+v_3P_{\lambda_n}+1}^{k_0} \dot{\eta}_t\left(\delta_4\left(\lambda_n + \frac{v_2}{\sqrt{n}}\right)\right)\lambda_n + op(1). \quad \textbf{(A.38)}$$

First, we can rewrite

$$\lambda_n' \sum_{t=k_0+v_3P_{\lambda_n}+1}^{k_0} \psi_t g(t/n) = \lambda_n' \sum_{t=k_0+v_3P_{\lambda_n}+1}^{k_0} \psi_t g(k_0/n) + \lambda_n' \sum_{t=k_0+v_3P_{\lambda_n}+1}^{k_0} \psi_t(g(t/n) - g(k_0/n)).$$

By Lemma 4, the first term converges weakly to $(\lambda_n' g(\tau_0)sg(\tau_0)'\lambda_n)^{1/2}W_1(-v_3P_{\lambda_n}) = W_1(-v_3)$, for $v_3 < 0$, where $P_{\lambda_n} = (\lambda_n' g(\tau_0)sg(\tau_0)'\lambda_n)^{-1}$ and $W_1(\cdot)$ is a Brownian motion process defined in $\mathbb{R}^+$ (for $v_3 > 0$, an equivalent result is obtained by symmetry). The second term is uniformly negligible because its variance will be an $op(1)$, given A5.2 and $P_{\lambda_n} = O(\|\lambda_n\|^{-2})$. Next, consider the second term of (A.38):

$$\frac{1}{2}\lambda_n' \sum_{t=k_0+v_3P_{\lambda_n}+1}^{k_0} \dot{\psi}(U_t + g(t/n)'\theta_n)g(t/n)g(t/n)$$

$$= \frac{1}{2}\lambda_n' \sum_{t=k_0+v_3P_{\lambda_n}+1}^{k_0} \dot{\psi}(U_t + g(t/n)'\theta_n)(g(t/n) - g(k_0/n))(g(t/n) - g(k_0/n))'\lambda_n \quad \textbf{(A.39)}$$

$$+ \frac{1}{2}\lambda_n' \sum_{t=k_0+v_3P_{\lambda_n}+1}^{k_0} \dot{\psi}(U_t + g(t/n)'\theta_n)g(k_0/n)g(k_0/n)'\lambda_n \quad \textbf{(A.40)}$$

$$+ \lambda_n' \sum_{t=k_0+v_3P_{\lambda_n}+1}^{k_0} \dot{\psi}(U_t + g(t/n)'\theta_n)(g(t/n) - g(k_0/n))g(k_0/n)'\lambda_n, \quad \textbf{(A.41)}$$

where $\theta_n = \delta_4(\lambda_n + v_2/\sqrt{n})$. Applying the MVT, with $x^* \in (t/n - k_0/n)$,

$$(\text{A.39}) = \frac{1}{2}\lambda_n' \sum_{t=k_0+v_3P_{\lambda_n}+1}^{k_0} \dot{\psi}(U_t + g(t/n)'\theta_n)\frac{dg(x^*)}{x}\frac{dg(x^*)}{x}'\left(\frac{t}{n} - \frac{k_0}{n}\right)^2\lambda_n,$$

which is upper bounded by

$$\left(\sup_x\left\|\frac{dg(x)}{x}\right\|\right)^2\left(\frac{k_0}{n} - \frac{k_0 + v_3P_{\lambda_n}}{n}\right)^2 v_3P_{\lambda_n}\sup_t|\dot{\psi}_t|\lambda_n'\lambda_n,$$

an $op(1)$ term, because $n\|\lambda_n\|^2 \to \infty$ and given that functions $g(\cdot)$ and $\dot{\psi}(\cdot)$ are bounded. Similar arguments show that the term (A.41) is also asymptotically negligible. Finally, Lemma 2 yields

$$(\text{A.40}) = \frac{1}{2}\lambda_n' v_3P_{\lambda_n}(m + op(1))g(\tau_0)g(\tau_0)'\lambda_n,$$

because $g(t/n)'\theta_n \leq \|g(t/n)\|\|\theta_n\| \to 0$, because $g(\cdot)$ is bounded. Combining these results and noting that $P_{\lambda_n} = (\lambda_n' g(\tau_0) s g(\tau_0)'\lambda_n)^{-1}$,

$$\text{(A.38)} \Rightarrow -\frac{1}{2}\lambda_n' v_3 P_{\lambda_n} m g(\tau_0) g(\tau_0)'\lambda_n - W_1(-v_3) = -\frac{m}{s} v_3 - W_1(-v_3).$$

Using similar arguments for $v_3 > 0$, we obtain that

$$\hat{v}_3 \Rightarrow \arg\min_v \left\{-W(v) + \frac{1}{2}\frac{m}{s}|v|\right\} = \arg\max_v \left\{W(v) - \frac{1}{2}\frac{m}{s}|v|\right\}.$$

Thus, with the change of variables $w = (m^2/s^2)v$ (and noting that $\hat{v}_3 = (\hat{k} - k_0)/P_{\lambda_n}$),

$$P_{\lambda_n}^{-1}(\hat{k} - k_0) \Rightarrow \frac{s^2}{m^2} \arg\max_w \left\{\left|\frac{s}{m}\right| W(w) + \frac{1}{2}\frac{m}{s}\frac{s^2}{m^2}|w|\right\} \stackrel{d}{=} \frac{s^2}{m^2} \arg\max_w \left\{W(w) + \frac{1}{2}|w|\right\},$$

and then, $(\lambda_n' g(\tau_0) g(\tau_0)'\lambda_n)(\hat{k} - k_0) \Rightarrow s m^{-2} \arg\max_w \{W(w) + \frac{1}{2}|w|\}$. ■

Lemmas

LEMMA 1. *Under A1.1, A1.2, A2.1–A2.5, and A2.7,* $\sup_{\tau\in\Pi}\sup_{\theta\in\Theta}\|n^{-1} \times \sum_{t=1}^{[n\tau]} \dot{\eta}_t(\theta) - \tau M(\theta)\| \xrightarrow{p} 0$, *where* $M(\theta)$ *is defined by A2.7.*

Proof. By the triangular inequality,

$$\sup_{\tau\in\Pi}\sup_{\theta\in\Theta}\left\|\frac{1}{n}\sum_{t=1}^{n\tau} \dot{\eta}_t(\theta) - \tau M(\theta)\right\| \leq \sup_{\tau\in\Pi}\sup_{\theta\in\Theta}\left\|\frac{1}{n}\sum_{t=1}^{[n\tau]} (\dot{\eta}_t(\theta) - E[\dot{\eta}_t(\theta)])\right\| \quad \textbf{(A.42)}$$

$$+ \sup_{\tau\in\Pi}\sup_{\theta\in\Theta}\left\|\frac{1}{n}\sum_{t=1}^{[n\tau]} E[\dot{\eta}_t(\theta)] - \tau M(\theta)\right\|. \quad \textbf{(A.43)}$$

By A2.7, (A.43) $\to 0$, whereas (A.42) $\xrightarrow{p} 0$ by A1.1 and A2.1–A2.5 applying Lemma A.3 of Andrews (1993). ■

LEMMA 2. *Let* $\Theta_0 \subset \Theta$ *be a compact subset of* $\mathbb{R}^p$, *containing neighborhoods of* θ_0. *Consider a sequence* $\{\theta_n, n \geq 1\} \in \Theta_0$ *such that* $\theta_n \xrightarrow[n\uparrow\infty]{} \theta_0$. *Then, under A1.1, A1.2, A2.1–A2.5, and A.27 it holds that*

$$\sup_{\tau\in\Pi}\left\|n^{-1}\sum_{t=1}^{[n\tau]} \dot{\eta}_t(\theta_n) - \tau M(\theta_0)\right\| \xrightarrow{p} 0.$$

Proof. By the triangle inequality, the left side of the preceding expression is bounded by

$$\sup_{\tau\in\Pi}\left\|\frac{1}{n}\sum_{t=1}^{[n\tau]} (\dot{\eta}_t(\theta_n) - E[\dot{\eta}_t(\theta_n)])\right\| + \sup_{\tau\in\Pi}\left\|\frac{1}{n}\sum_{t=1}^{[n\tau]} (E[\dot{\eta}_t(\theta_n)] - E[\dot{\eta}_t(\theta_0)])\right\|$$

$$+ \sup_{\tau\in\Pi}\left\|\frac{1}{n}\sum_{t=1}^{[n\tau]} E[\dot{\eta}_t(\theta_0)] - \tau M(\theta_0)\right\| = (I) + (II) + (III).$$

By A2.7, $(III) \to 0$, whereas $(I) \xrightarrow{p} 0$, by A1.1 and A2.1–A2.5, because

$$(I) \leq \sup_{\tau \in \Pi} \sup_{\theta \in \Theta_0} \left\| \frac{1}{n} \sum_{t=1}^{[n\tau]} (\dot{\eta}_t(\theta) - E[\dot{\eta}_t(\theta)]) \right\| \xrightarrow{p} 0,$$

by Lemma A.3 of Andrews (1993). It remains to study (II): (i) by the tightness condition of $\{F_n, n \geq 1\}$, we obtain that $n^{-1} \sum_{t=1}^{n} P(Z_t \notin C_j) \to 0$ as $j \to \infty$, for some sequence of compact sets $\{C_j, j \geq 1\}$ in Z, and (ii) $\forall j \geq 1$,

$$\sup_{n \geq 1} \sup_{\tau \in \Pi} \left\| \frac{1}{n} \sum_{t=1}^{[n\tau]} E[\dot{\eta}_t(\theta_n) - \dot{\eta}_t(\theta_0)] I(Z_t \in C_j) \right\|$$

$$\leq \sup_{z \in C_j} \|\dot{\eta}(z, \theta_n) - \dot{\eta}(z, \theta_0)\| \to 0 \quad \text{for } \theta_n \to \theta_0,$$

for a function $\dot{\eta}(\cdot)$ defined in $(z, \theta) \in Z \times \Theta$, continuous by A1.1, and thus uniformly continuous in the compact set C_j. (iii) By results (i) and (ii),

$$\sup_{\tau \in \Pi} \left\| \frac{1}{n} \sum_{t=1}^{[n\tau]} E[\dot{\eta}_t(\theta_n) - \dot{\eta}_t(\theta_0)] \right\| \to 0, \text{ as } \theta_n \to \theta_0. \qquad \blacksquare$$

For convenience in the subsequent discussion, we provide the following definition.

DEFINITION 2. *On a probability space* $(\Omega, \mathcal{F}, P)$, *the sequence of pairs* $\{Y_t, \mathcal{F}_t\}_0^\infty$, *where* $\mathcal{F}_t$ *is an increasing sequence of* σ*-subfields of* $\mathcal{F}$ *and the* Y_t *are integrable random variables, is an* L_r *mixingale if, for* $r \geq 1$, *there exist sequences of nonnegative constants* $\{c_t\}_0^\infty$ *and* $\{\zeta_m\}_0^\infty$ *such that* $\zeta_m \to 0$ *as* $m \to 0$, *and*

(i) $\|E(Y_t | \mathcal{F}_{t-m})\|_r \leq c_t \zeta_m$,
(ii) $\|Y_t - E(Y_t | \mathcal{F}_{t+m})\|_r \leq c_t \zeta_{m+1}$,

hold for all $t \geq 0$ *and* $m \geq 0$. *When* $\zeta_m = O(m^{-q})$ *for all* $q > q_0$, *we say that* $\{\zeta_m\}$ *is of size* $-q_0$.

Define $S_j = \sum_1^j Y_t$ with $S_0 = 0$ and construct the random function $v_n(\tau) = n^{-1/2} S_{[n\tau]}$. Note that $\{v_n(\tau), n \geq 1\}$ belongs to the bounded cadlag function space in $\mathbb{R}^p$ and is defined on $\Pi \subset [0,1]$. The following lemma generalizes the Hájek and Rényi inequality to L_r mixingales for $r > 1$.

LEMMA 3. *Let* $\{a_k\}_{k \geq 1}$ *be a sequence of decreasing positive constants.*

(3.1) If $\{Y_t, \mathcal{F}_t\}$ *is a* L_r *mixingale with constants* $\{c_t\}$ *such that* $\sup_{t \leq n} \|Y_t\|_r < \infty$ *and*
(i) $1 < r < 2$ *with* $\{\zeta_m\}$ *of size* -1,
(ii) $r = 2$ *with* $\{\zeta_m\}$ *of size* $-\frac{1}{2}$,
or
(iii) $r > 1$ *with* $\sum_{k=1}^\infty \zeta_k < \infty$ *and* $\sup_{t \leq n} \|Y_t\|_2 < \infty$.

Then there exists a $K < \infty$ such that

$$\Pr\left\{\max_{m\le k\le n} a_k|S_k| > \varepsilon\right\} \le \varepsilon^{-p}K\left(ma_m^p + \sum_{j=m+1}^{n} a_j^p\right), \tag{A.44}$$

for every $\varepsilon > 0$ and $m > 0$, $p = r$ under (i) or (ii) and $p = 2$ under (iii).

(3.2) *If $Y_t \equiv \eta_t$ and $\{\eta_t, \mathcal{F}_t\}$ satisfies Assumption A2.4, there exists a $K < \infty$ such that*

$$\Pr\left\{\max_{m\le k\le n} \sup_{\substack{\alpha\in\mathbb{R}^p \\ \|\alpha\|=1}} a_k|\alpha' S_k| > \varepsilon\right\} \le \varepsilon^{-2}K\left(ma_m^2 + \sum_{j=m+1}^{n} a_j^2\right), \tag{A.45}$$

for every $\varepsilon > 0$ and $m > 0$.

(3.3) *For $a_k = k^{-1}$, it holds that both probabilities on the left side of (A.44) and (A.45) above are $O(m^{-1}\varepsilon^{-2})$.*

Proof of (3.1). By the triangle inequality,

$$\Pr\left\{\max_{m\le k\le n} a_k|S_k| > \varepsilon\right\} \le \Pr\left\{\max_{m\le k\le n} a_k|S_m| > \frac{\varepsilon}{2}\right\} + \Pr\left\{\max_{m\le k\le n} a_k|S_k - S_m| > \frac{\varepsilon}{2}\right\}. \tag{A.46}$$

With $\{a_k\}$ decreasing, we have that $\max_{m\le k\le n} a_k = a_m$, and the first term of the right side will be upper bounded by

$$\Pr\left\{|S_m| > \frac{\varepsilon}{2a_m}\right\} \le \varepsilon^{-p}\|S_m\|_p^p 2^p a_m^p \le \varepsilon^{-p} m \sup_{t\le n}\|Y_t\|_p^p 2^p a_m^p = \varepsilon^{-p}K_1 m a_m^p. \tag{A.47}$$

The two inequalities are Markov's and Minkowski's, with p already defined in the lemma according to the cases. The last equality sets $K_1 = 2^p \sup_{t\le n}\|Y_t\|_p^p$. The second term on the right side of (A.46) will be bounded by

$$\Pr\left\{\max_{m\le k\le n}\left|\sum_{t=m+1}^{k} a_t Y_t\right| > \frac{\varepsilon}{4}\right\} + \Pr\left\{\max_{m\le k\le n}\left|\sum_{t=m+1}^{k} (a_t - a_k) Y_t\right| > \frac{\varepsilon}{4}\right\}$$

$$\le \left(\frac{4}{\varepsilon}\right)^p \left(\left\|\max_{m\le k\le n}\left|\sum_{t=m+1}^{k} a_t Y_t\right|\right\|_p^p + \left\|\max_{m\le k\le n}\left|\sum_{t=m+1}^{k} (a_t - a_k) Y_t\right|\right\|_p^p\right), \tag{A.48}$$

by Markov's inequality. Let $q_{1t} \equiv a_t Y_t$, $\forall t$. Then, we have that $\{q_{1t}, \mathcal{F}_t\}$ is a mixingale with constants $c_{1t} = c_t a_t$ and the first summand of (A.48) will be bounded by

$$\left(\frac{4}{\varepsilon}\right)^p K_2 \sum_{t=m+1}^{k} c_{1t}^p \le \left(\frac{4}{\varepsilon}\right)^p K_2 \sup_{t\le n}|c_t|^p \sum_{t=m+1}^{k} a_t^p, \tag{A.49}$$

for $K_2 < \infty$. With $p = r$, the first bound is due to Davidson (1994, Theorem 16.11) for (i) and to McLeish (1975a, Theorem 1.6) for (ii), and, with $p = 2$, it is due to Hansen (1991, Lemma 2) for (iii). Similar arguments apply for the second term of (A.48), which will be upper bounded by

$$\left(\frac{4}{\varepsilon}\right)^p K_3 \sum_{t=m+1}^{k} c_{2t}^p \leq \left(\frac{4}{\varepsilon}\right)^p K_3 \sum_{t=m+1}^{k} c_{1t}^p \tag{A.50}$$

for $K_3 < \infty$ and $c_{2t} = c_t(a_t - a_k)$. The last inequality is obtained given that $a_t \geq a_k \geq 0$, for all $t \leq k$. The result of (3.1) is established with $K = 4^p \max\{K_1, \max\{K_2, K_3\} \sup_{t \leq n} |c_t|^p\}$. ■

Proof of (3.2). We verify the conditions of (3.1) for this particular case. Because $\{\eta_t\}$ is L_2-*NED* with respect to $\{w_t\}$ of size $-\frac{1}{2}$ with $d_t = 1$, it follows that

$$\sup_{\substack{\alpha \in \mathbb{R}^p \\ \|\alpha\|=1}} \|\alpha' \eta_t\|_2 = \sup_{\substack{\alpha \in \mathbb{R}^p \\ \|\alpha\|=1}} \left(E\left|\sum_{i=1}^{p} \alpha_i \eta_{ti}\right|^2\right)^{1/2} \leq \sup_{\substack{\alpha \in \mathbb{R}^p \\ \|\alpha\|=1}} \left(E\left[\sum_{i=1}^{p} \alpha_i^2 \sum_{i=1}^{p} \eta_{ti}^2\right]\right)^{1/2} = \|\eta_t\|_2,$$

by the Cauchy–Schwarz inequality, and similarly,

$$\sup_{\substack{\alpha \in \mathbb{R}^p \\ \|\alpha\|=1}} \|\alpha' \eta_t - E_{t-m}^{t+m} \alpha' \eta_t\|_2 \leq \sup_{\substack{\alpha \in \mathbb{R}^p \\ \|\alpha\|=1}} \left(E\left[\sum_{i=1}^{p} \alpha_i^2 \sum_{i=1}^{p} (\eta_{ti} - E_{t-m}^{t+m} \eta_{ti})^2\right]\right)^{1/2}$$

$$= \|\eta_t - E_{t-m}^{t+m} \eta_t\|_2 \leq d_t v_m.$$

Therefore, $\{\alpha' \eta_t\}$ is also L_2-*NED* of size $-\frac{1}{2}$ and $d_t = 1$ (observe that if $\|\alpha\| \neq 1$, $d_t = \sup_{\alpha \in \mathbb{R}^p} \|\alpha\| < \infty$, $\forall n, t$). Furthermore, under Assumption A2.4, $\{\alpha' \eta_t\}$ is an L_2 mixingale of $-\frac{1}{2}$ and constants $c_t = d_t \max_{t \leq n}\{1, \|\eta_t\|_2\}$ (see, e.g., Wooldridge and White, 1988, Proposition 2.9). Given that $\sup_{t \leq n} \|\eta_t\|_2 < \infty$, conditions of (3.1)(ii) are fulfilled.

Proof of (3.3). For $a_k = k^{-1}$, the result follows given that $\sum_{k=m}^{\infty} k^{-2} = O(m^{-1})$. ■

LEMMA 4. *For $Y_t \equiv \eta_t$, under A1.1, A2.1–A2.4, and A2.8, $v_n(\tau) \Rightarrow S^{1/2} B(\tau)$, where $\Rightarrow$ denotes weak convergence, in the $D([0,1])$ space equipped with the Skorokhod metric. Here $B(\cdot)$ is a p-vector of independent Brownian motion processes. The process can be defined in $C([0,1])$, for which there exists an equivalence between the Skorokhod and uniform metrics.*

Proof. To prove this lemma we use the results of Lemma A.4 of Andrews (1993), with the difference that, in this case, the triangular arrays he used are constant across the subscript n. Then, we have to prove that the process $\{v_n(\tau), n \geq 1\}$ is such that

(i) $\forall \alpha \in \mathbb{R}^p, \alpha' v_n(\tau) \Rightarrow \alpha' S^{1/2} B(\tau)$,
(ii) $\{v_n(\tau), n \geq 1\}$ has asymptotically independent increments.

To obtain (i), under A2.1, A2.2, and A2.8, we apply Corollary 3.2 of Wooldridge and White (1988), which utilizes the results of McLeish (1977). Noting that the mentioned corollary yields weak convergence of the standard partial sum process in $D([0,1])$ with the Skorokhod metric and the σ-field generated by it, this can be converted into weak convergence in $D([0,1])$ with the uniform metric and the σ-field generated by the closed balls under the uniform metric. This is treated by Andrews (1993). To obtain (ii) it will be enough to prove that

$$\begin{pmatrix} v_n(\tau_2) - v_n(\tau_1) \\ v_n(\tau_0) \end{pmatrix} \xrightarrow{d} N\left(0, \begin{pmatrix} (\tau_2 - \tau_1) S & 0 \\ 0 & \tau_0 S \end{pmatrix}\right), \qquad \forall 0 \leq \tau_0 < \tau_1 \leq \tau_2 \leq 1.$$

The Cramér–Wold device leads to

$$\alpha_1' v_n(\tau_2) - \alpha_1' v_n(\tau_1) + \alpha_2' v_n(\tau_0) \xrightarrow{d} N(0, (\tau_2 - \tau_1)\alpha_1' S\alpha_1 + \tau_0 \alpha_2' S\alpha_2), \quad \forall \alpha_1, \alpha_2 \in \mathbb{R}^p.$$

To obtain the preceding result, we use again Corollary 3.2 of Wooldridge and White (1988),

$$\begin{aligned}\alpha_1' v_n(\tau_2) - \alpha_1' v_n(\tau_1) + \alpha_2' v_n(\tau_0) &= \alpha_1' \frac{1}{\sqrt{n}} \sum_{t=[n\tau_1]+1}^{[n\tau_2]} \eta_t + \alpha_2' \frac{1}{\sqrt{n}} \sum_{t=1}^{[n\tau_0]} \eta_t \\ &\Rightarrow \alpha_1' S^{1/2} B(\tau_2 - \tau_1) + \alpha_2' S^{1/2} B(\tau_0) \\ &\stackrel{d}{=} N(0, (\tau_2 - \tau_1)\alpha_1' S\alpha_1 + \tau_0 \alpha_2' S\alpha_2).\end{aligned}$$

And we get the desired result. ■

LEMMA 5. *For $Y_t \equiv \eta_t$, under Assumptions A1.1, A2.1–A2.4, and A2.8, there exists a $K < \infty$ such that, for every $\varepsilon > 0$ and large n,* $\Pr\{\sup_{\tau \le 1} \sup_{\alpha \in \mathbb{R}^p, \|\alpha\|=1} \alpha' |v_n(\tau)| > \varepsilon\} \le \varepsilon^{-2} K$.

Proof. By Lemma 4 and the continuous mapping theorem, $\sup_{\alpha \in \mathbb{R}^p, \|\alpha\|=1} \alpha' v_n(\tau) \Rightarrow \sup_{\alpha \in \mathbb{R}^p, \|\alpha\|=1} (\alpha' S\alpha)^{1/2} B(\tau)$. Let $Y = \sup_\tau |B(\tau)|$, a process with finite second moment because $P\{\sup_\tau B(\tau) \le d\} = 2\Phi(d)$, $d \ge 0$, and $B(\cdot)$ is symmetric around zero. Thus, the Markov inequality leads to

$$\Pr\left\{\sup_{\tau \le 1} \sup_{\substack{\alpha \in \mathbb{R}^p \\ \|\alpha\|=1}} \alpha' |v_n(\tau)| > \varepsilon\right\} \to \Pr\left\{Y > \frac{\varepsilon}{\sup_{\substack{\alpha \in \mathbb{R}^p \\ \|\alpha\|=1}} (\alpha' S\alpha)^{1/2}}\right\} \le \frac{E(Y^2)}{\varepsilon^2} \sup_{\substack{\alpha \in \mathbb{R}^p \\ \|\alpha\|=1}} (\alpha' S\alpha).$$

The Cauchy–Schwarz inequality establishes that $(\alpha' S\alpha) = \|S^{1/2}\alpha\| \le \|S^{1/2}\|\|\alpha\|$ and S is a finite matrix by A2.8. The result follows with $K = E(Y^2)\|S^{1/2}\| < \infty$. ■

LEMMA 6. *For $Y_t \equiv \eta_t$, under Assumptions A1.1, A2.1–A2.4, and A2.8, there exists a $K < \infty$ such that, for every $\varepsilon > 0$ and large n,* $\Pr\{\sup_{\alpha \in \mathbb{R}^p, \|\alpha\|=1} \alpha' |v_n(1) - v_n(\tau_0)| > \varepsilon\} \le \varepsilon^{-2} K$.

Proof. Lemma 4 yields $\alpha'(v_n(1) - v_n(\tau_0)) \xrightarrow{d} \alpha' S^{1/2} B(1 - \tau_0) \stackrel{d}{=} N(0, (1 - \tau_0)\alpha' S\alpha)$. Thus,

$$\begin{aligned}\Pr\left\{\sup_{\substack{\alpha \in \mathbb{R}^p \\ \|\alpha\|=1}} \alpha' |v_n(1) - v_n(\tau_0)| > \varepsilon\right\} &\to \Pr\left\{|Z| > \frac{\varepsilon}{(1 - \tau_0) \sup_{\substack{\alpha \in \mathbb{R}^p \\ \|\alpha\|=1}} (\alpha' S\alpha)^{1/2}}\right\} \\ &\le \frac{E(Z^2)}{\varepsilon^2} (1 - \tau_0) \sup_{\substack{\alpha \in \mathbb{R}^p \\ \|\alpha\|=1}} (\alpha' S\alpha),\end{aligned}$$

by Markov inequality, where Z represents the standard normal variable and hence with a second moment equal to one. The result follows with $K = (1 - \tau_0)\|S^{1/2}\| < \infty$. ■